



US009441205B2

(12) **United States Patent**
Joshi(10) **Patent No.:** **US 9,441,205 B2**
(45) **Date of Patent:** **Sep. 13, 2016**(54) **TWO PLASMID MAMMALIAN EXPRESSION SYSTEM**(76) Inventor: **Vishwas Joshi**, Maharashtra (IN)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/124,383**(22) PCT Filed: **Jun. 8, 2012**(86) PCT No.: **PCT/IN2012/000405**

§ 371 (c)(1),

(2), (4) Date: **Feb. 27, 2014**(87) PCT Pub. No.: **WO2013/046216**PCT Pub. Date: **Apr. 4, 2013**(65) **Prior Publication Data**

US 2014/0212922 A1 Jul. 31, 2014

(30) **Foreign Application Priority Data**

Jun. 8, 2011 (IN) 1679/MUM/2011

(51) **Int. Cl.****C12N 15/85** (2006.01)**C12N 7/00** (2006.01)**C12N 15/86** (2006.01)(52) **U.S. Cl.**CPC **C12N 7/00** (2013.01); **C12N 15/85** (2013.01); **C12N 15/86** (2013.01); **C12N 2760/00051** (2013.01); **C12N 2760/18451** (2013.01); **C12N 2800/40** (2013.01); **C12N 2840/203** (2013.01)(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,716,821 A 2/1998 Wertz et al.
2008/0274130 A1 11/2008 Rupprecht et al.
2010/0028377 A1* 2/2010 Jin et al. 424/199.1

FOREIGN PATENT DOCUMENTS

WO WO 97/06270 2/1997
WO WO 98/13501 4/1998
WO WO 00/26517 5/2000
WO WO 2004/113517 A2 12/2004

OTHER PUBLICATIONS

Showronek et al., Analytical Biochemistry, 2002, vol. 300, pp. 185-191.*
Martin et al., Journal of Virology, 2006, vol. 80, pp. 5708-5715.*
Radcliffe et al., GENE Therapy, 2004, vol. 11, pp. 1673-1674.*
International Search Report for PCT/IN2012/000405 mailed Mar. 4, 2013.
Martin et al., "RNA Polymerase II-Controlled Expression of Antigenomic RNA Enhances the Rescue Efficacies of Two DifferentMembers of the *Mononegavirales* Independently of the Site of Viral Genome Replication," Journal of Virology, Jun. 2006, vol. 80, No. 12, 5708-5715.

Groseth et al., "RNA Polymerase I-Driven Minigenome System for Ebola Viruses," Journal of Virology, Apr. 2005, vol. 79, No. 7, 4425-4433.

Yanai et al., "Development of a novel Borna disease virus reverse genetics system using RNA polymerase II promoter and SV40 nuclear import signal," Microbes and Infection 8 (2006) 1522-1529.
Finke et al., "Differential Transcription Attenuation of Rabies Virus Genes by Intergenic Regions: Generation of Recombinant Viruses Overexpressing the Polymerase Gene," Journal of Virology, Aug. 2000, vol. 74, No. 16, 7261-7269.

Szymczak et al., "Development of 2A peptide-based strategies in the design of multicistronic vectors," Expert Opin Biol Ther, (2005) 5(5) 627-638.

Szymczak et al., "Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector," Nature Biotechnology vol. 22, No. 5, May 2004, 589-594.

Huang et al., "Development of a reverse genetics system for a human rabies virus vaccine strain employed in China," Virus Research 149 (2010) 28-35.

Parida et al., "Rescue of a chimeric rinderpest virus with the nucleocapsid protein derived from peste-des-petits-ruminants virus: use as a marker vaccine," Journal of General Virology (2007), 88, 2019-2027.

Bailey et al., "Reverse genetics for peste-des-petits-ruminants virus (PPRV): Promoter and protein specificities," Virus Research 126 (2007) 250-255.

Raha et al., "Development of a reconstitution system for Rinderpest virus RNA synthesis in vitro," Virus Research 99 (2004) 131-138. Supplementary European Search Report and Written Opinion issued in EP 12 83 5226 dated Jan. 14, 2015.

* cited by examiner

Primary Examiner — Michele K Joike

Assistant Examiner — Mindy G Brown

(74) Attorney, Agent, or Firm — RatnerPrestia

(57) **ABSTRACT**

Reverse engineering has offered new ways of studying the pathology of RNA viral infections, new more efficient devices of synthesizing recombinant viruses and developing vaccines and also demonstrated the versatility and efficiency of RNA dependent RNA polymerase RDRP system as an expression system. However, the currently used methods require a repertoire of complex, difficult-to-use tools. Present invention describes, a simpler plasmid based mammalian expression system that uses the RDRP enzyme activity for expression of recombinant proteins or RNA from viral minigenomes and rescue of recombinant viruses from cDNAs encoding entire genome(s) of negative stranded RNA viruses. This system will be useful for expression of recombinant proteins, therapeutic RNA molecules including anti-sense and/or selecting interfering RNA and Ribozymes. This system can also be used for gene therapy and producing recombinant viruses for production of new vaccines.

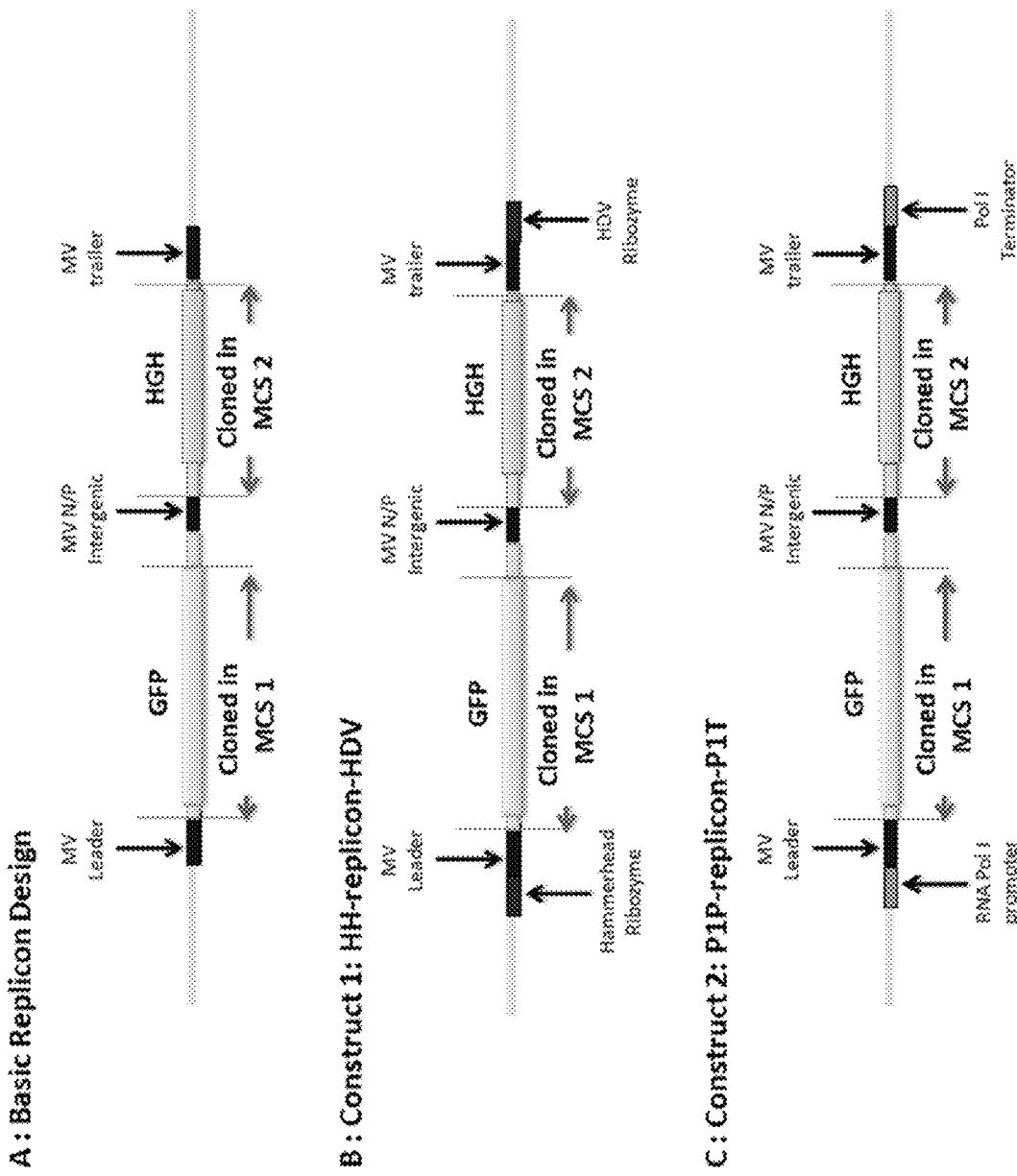


Fig 1

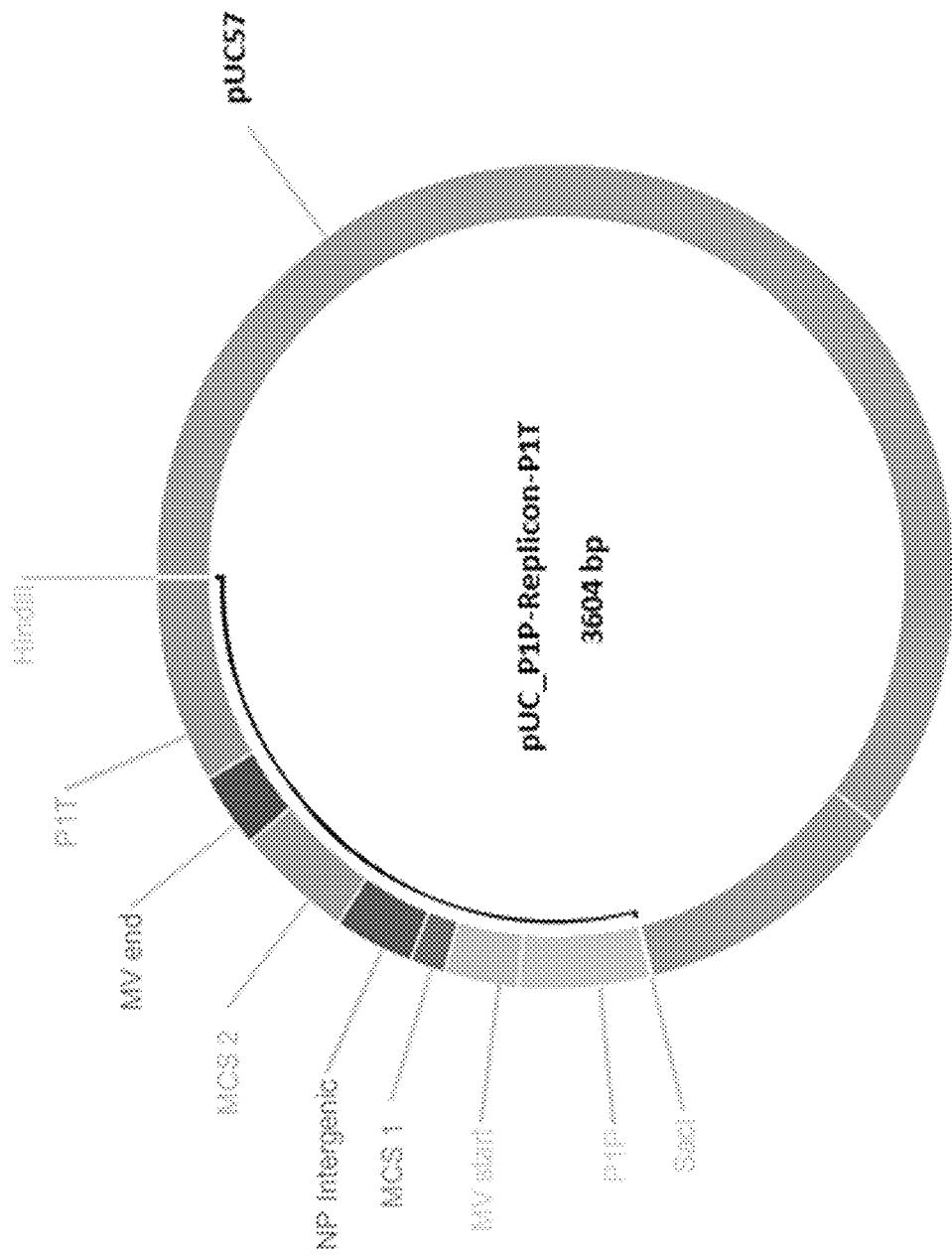


Fig 2A cloning plasmid 1

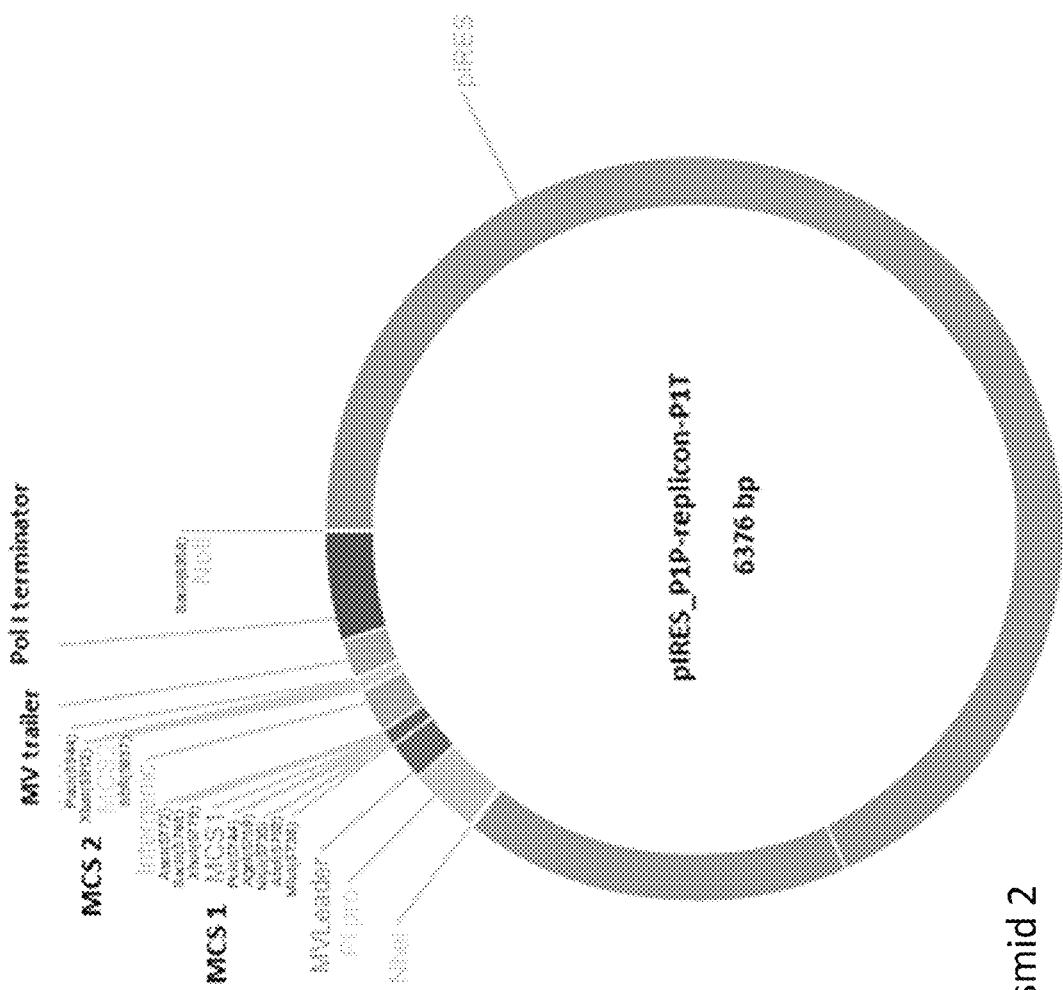


Fig 2B cloning plasmid 2

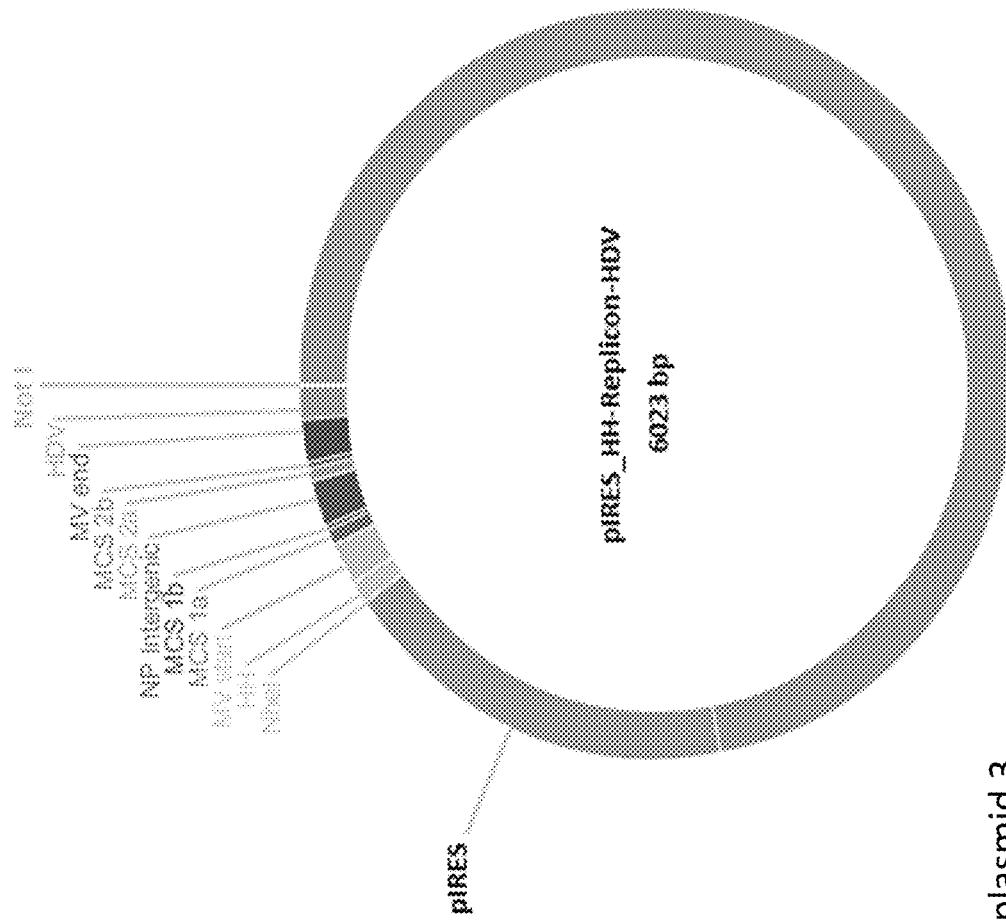


Fig 2C cloning plasmid 3

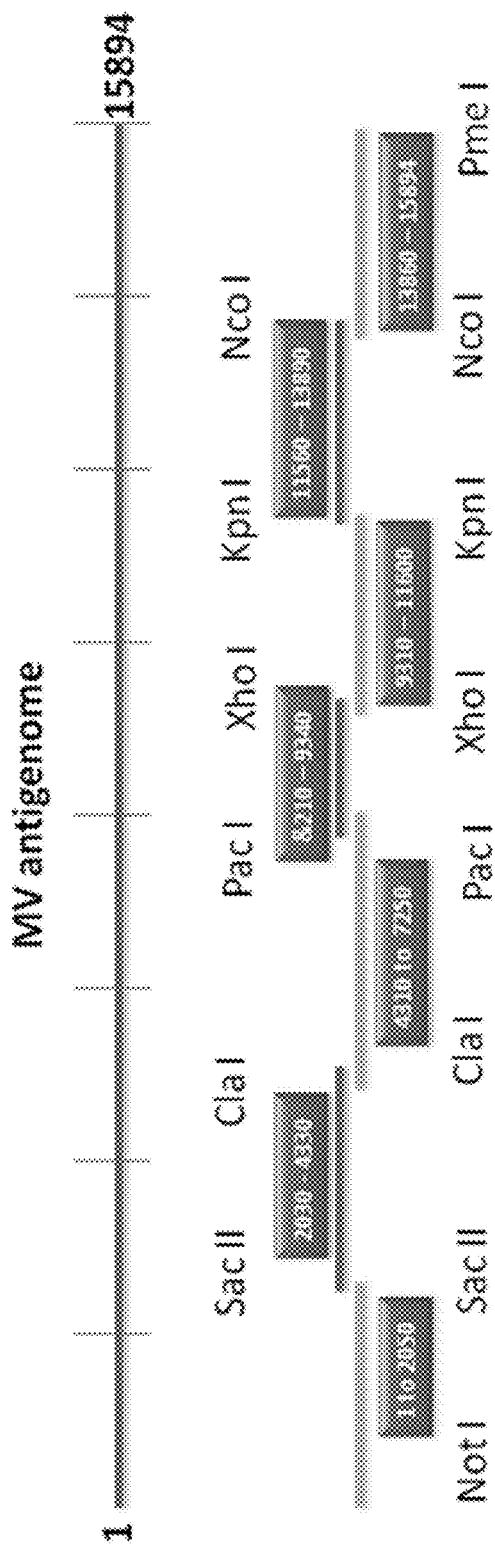


Fig 3A

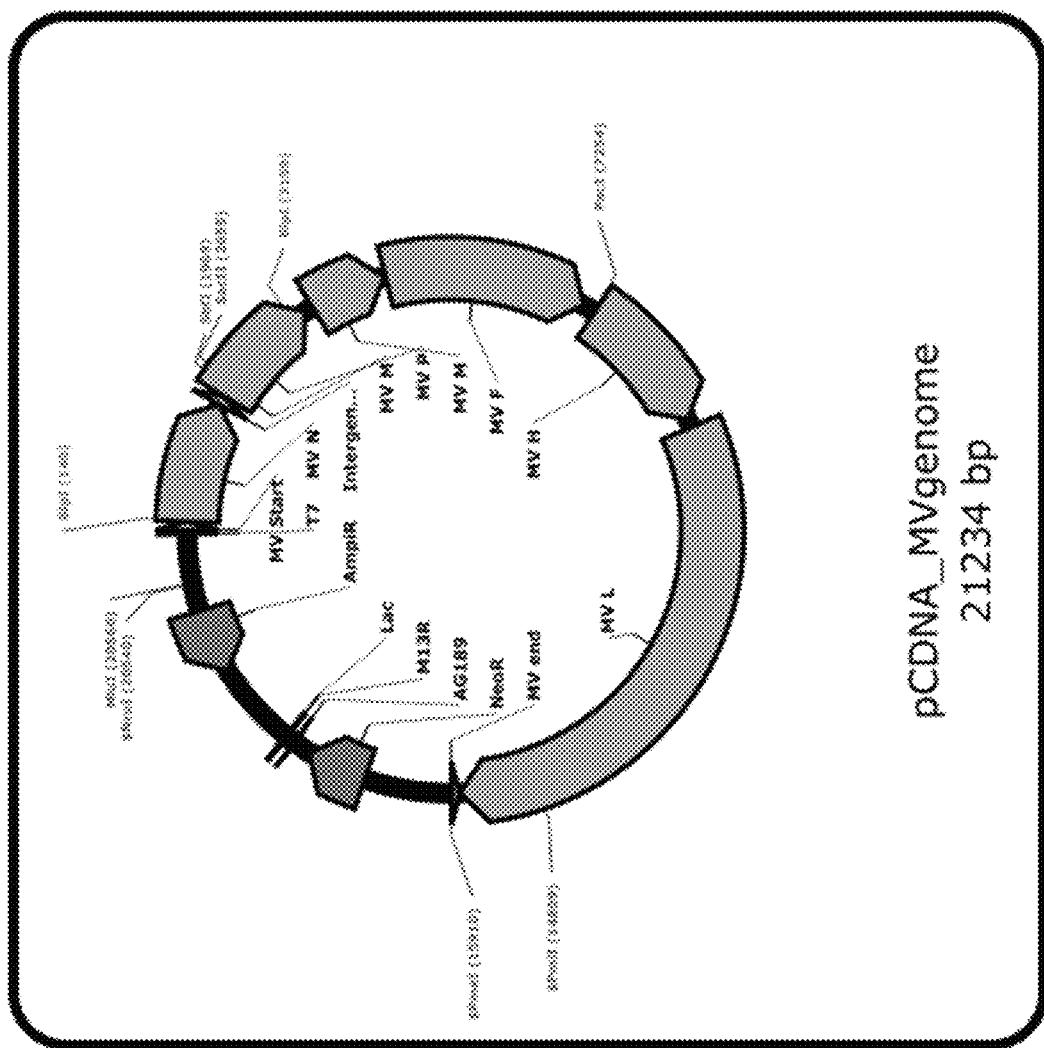


Fig 3B

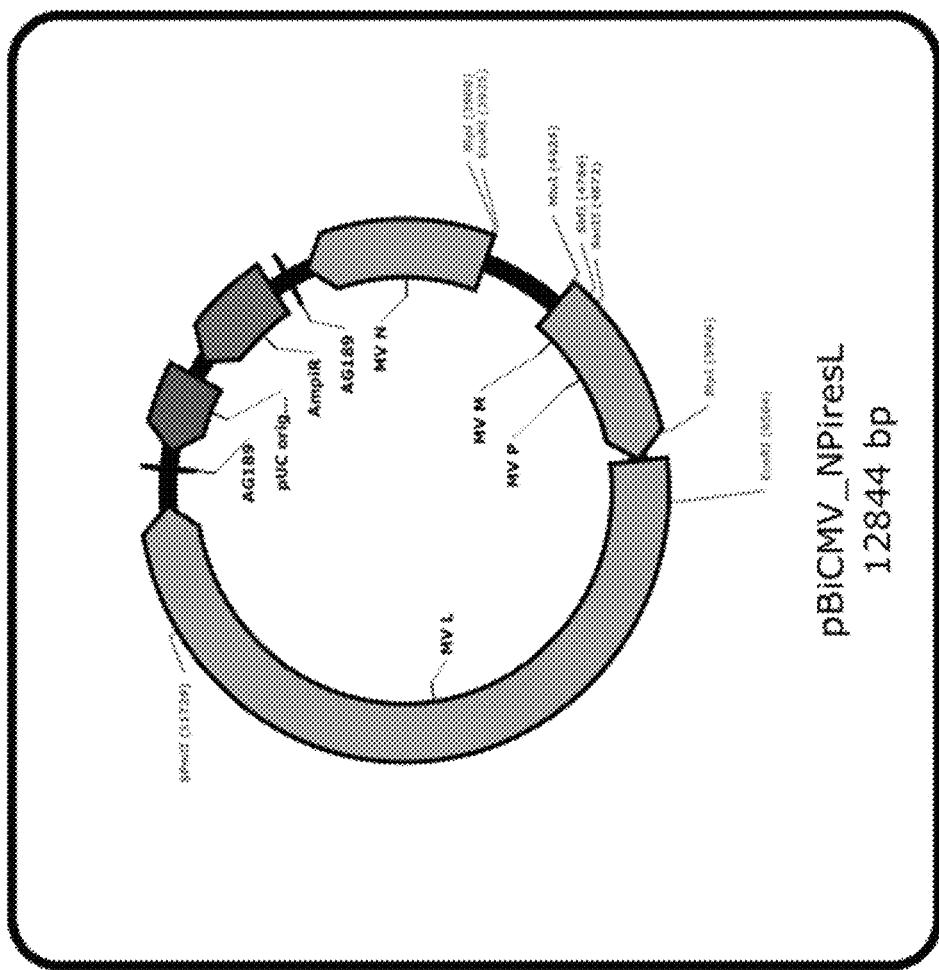
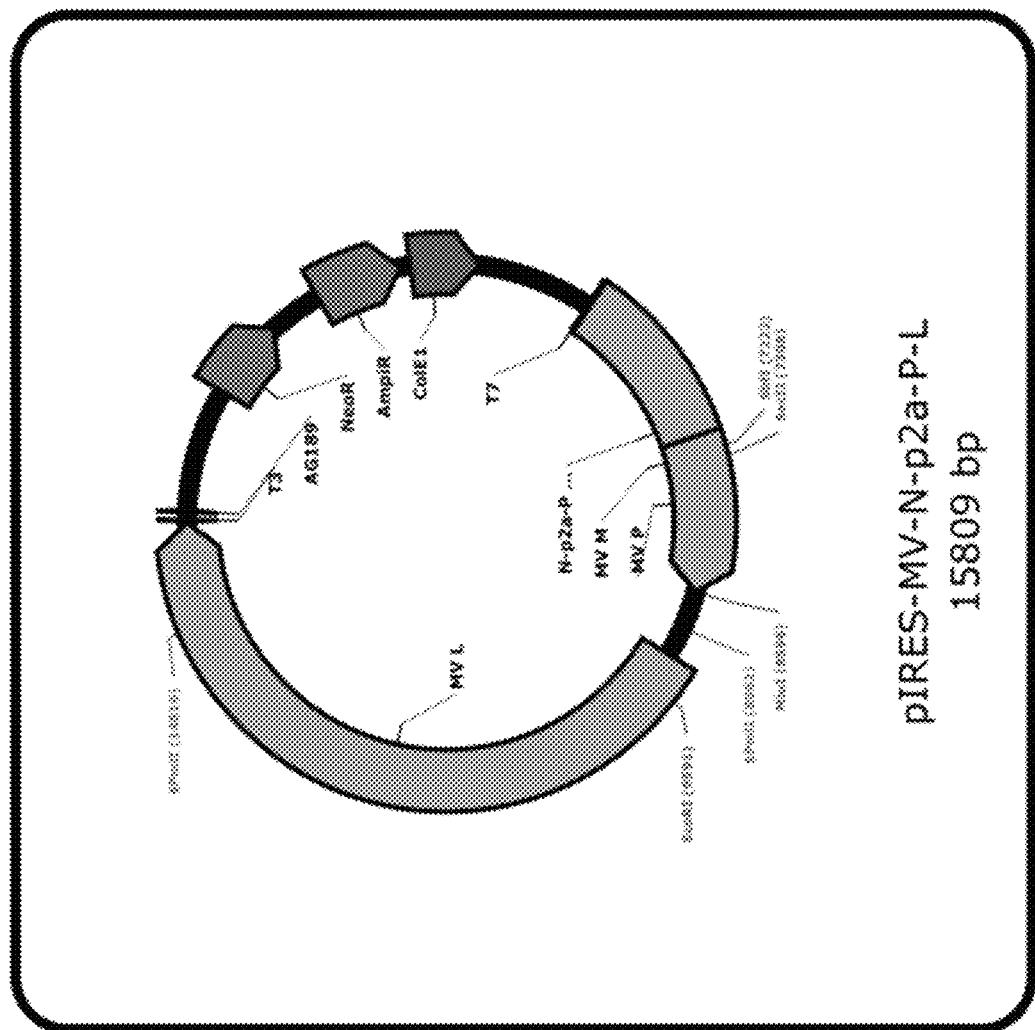


Fig 4A



pIRESV-MV-N-p2a-P-L
15809 bp

Fig 4B

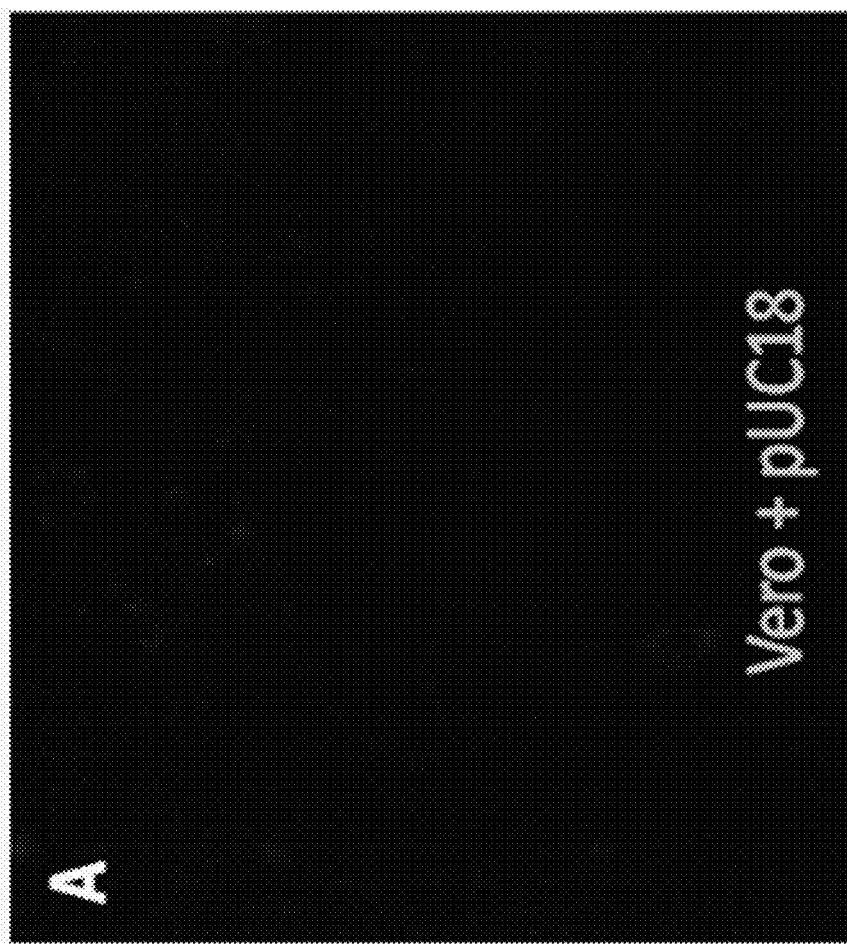


Fig 5A

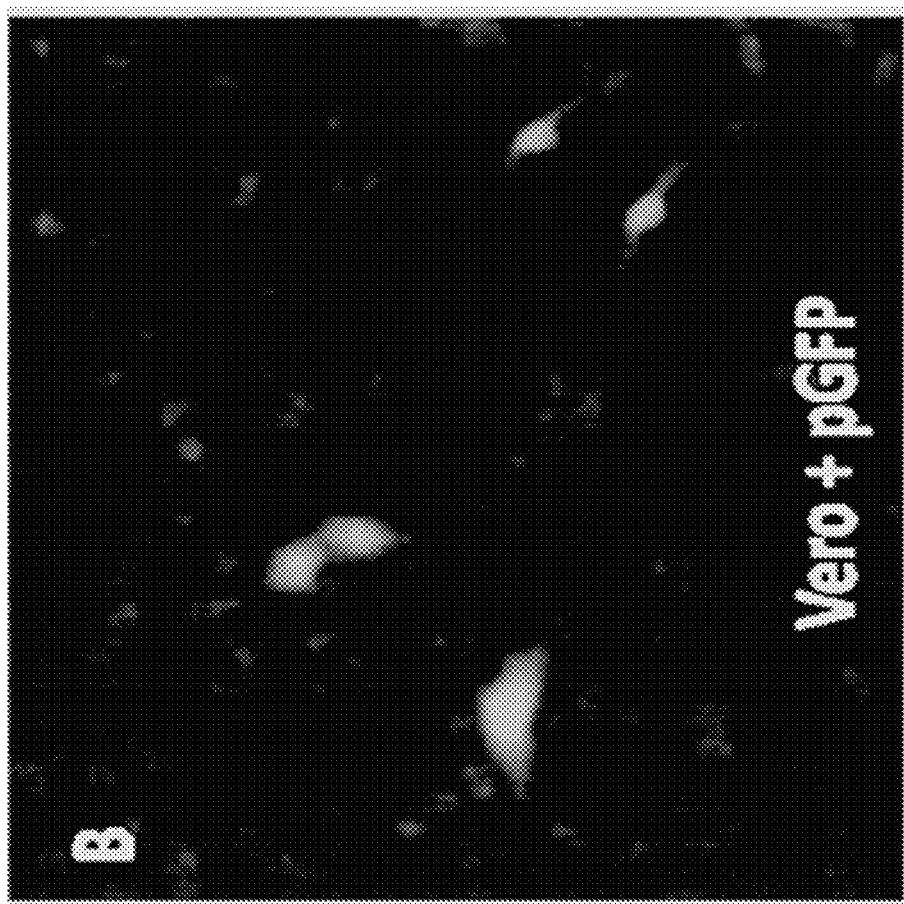


Fig 5B

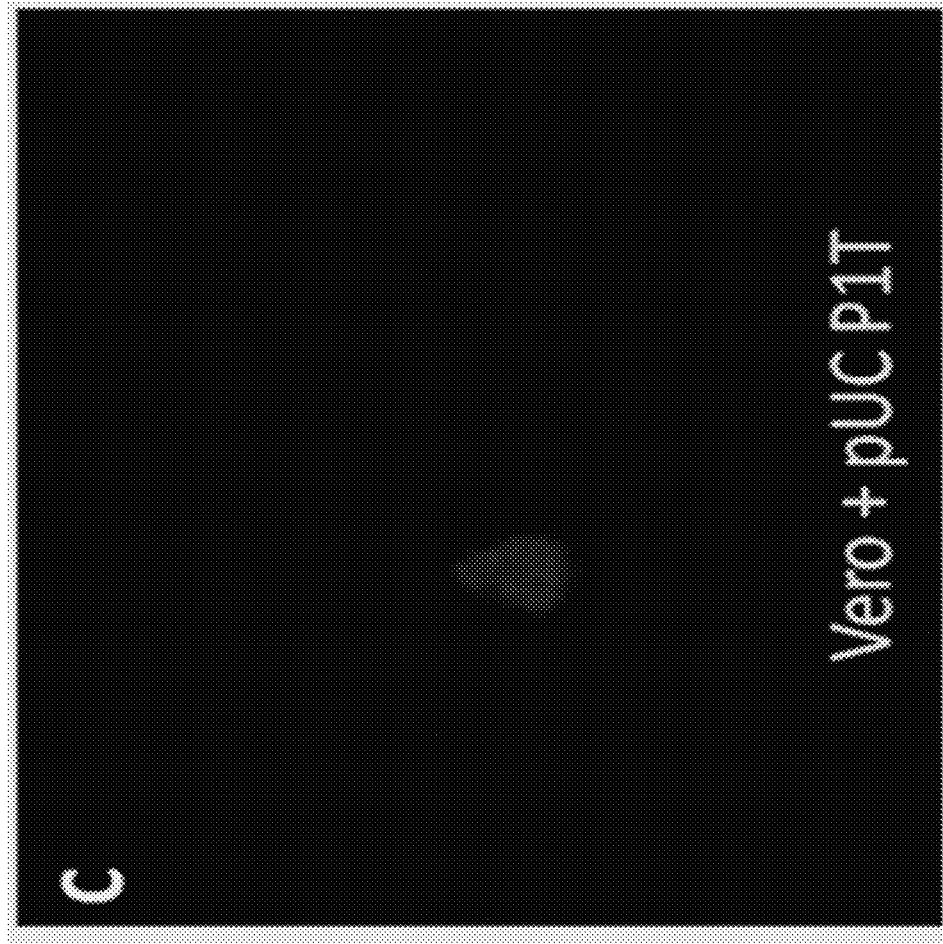
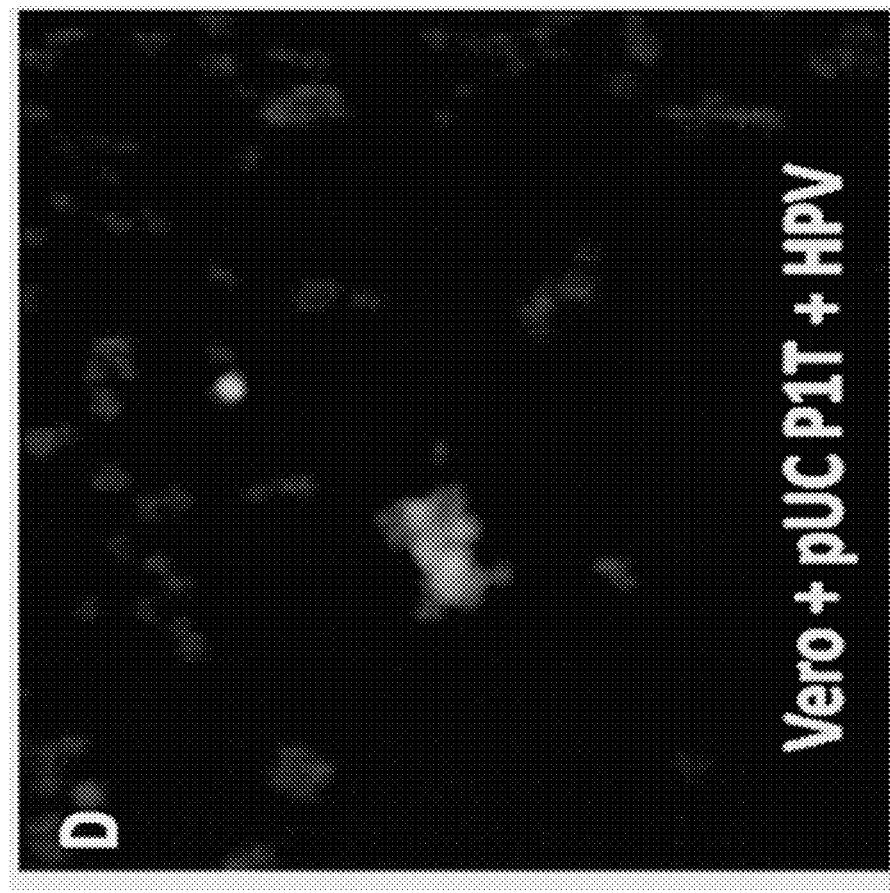


Fig 5C



Vero + pUC PlT + HPV

Fig 5D



Fig 5E

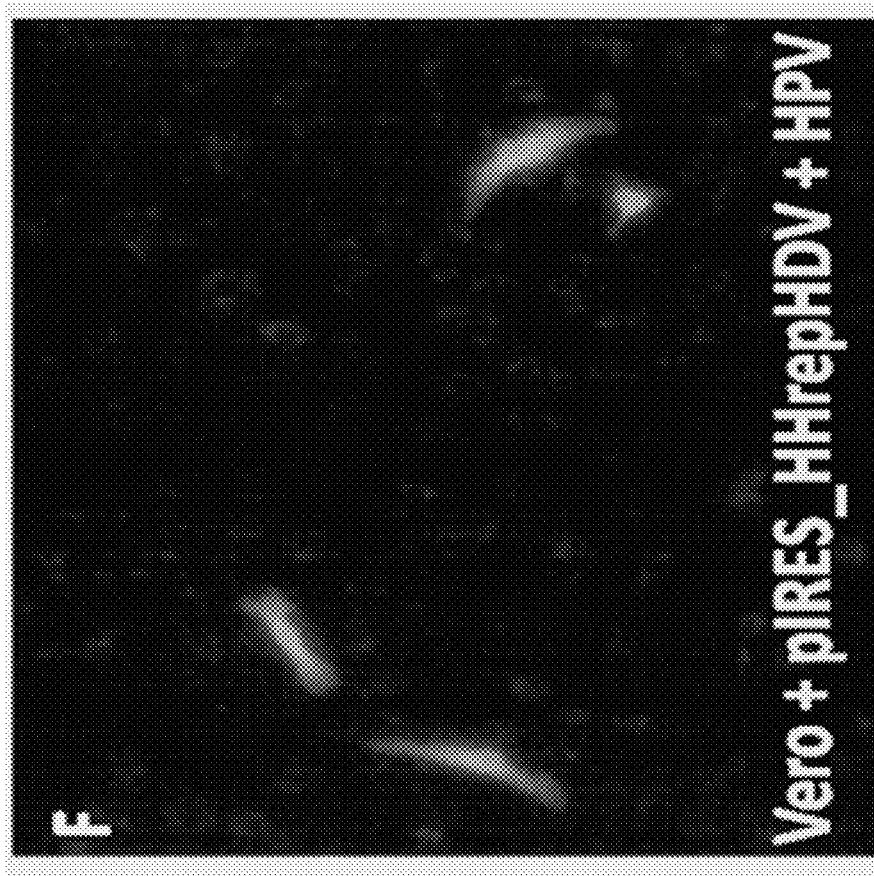


Fig 5F

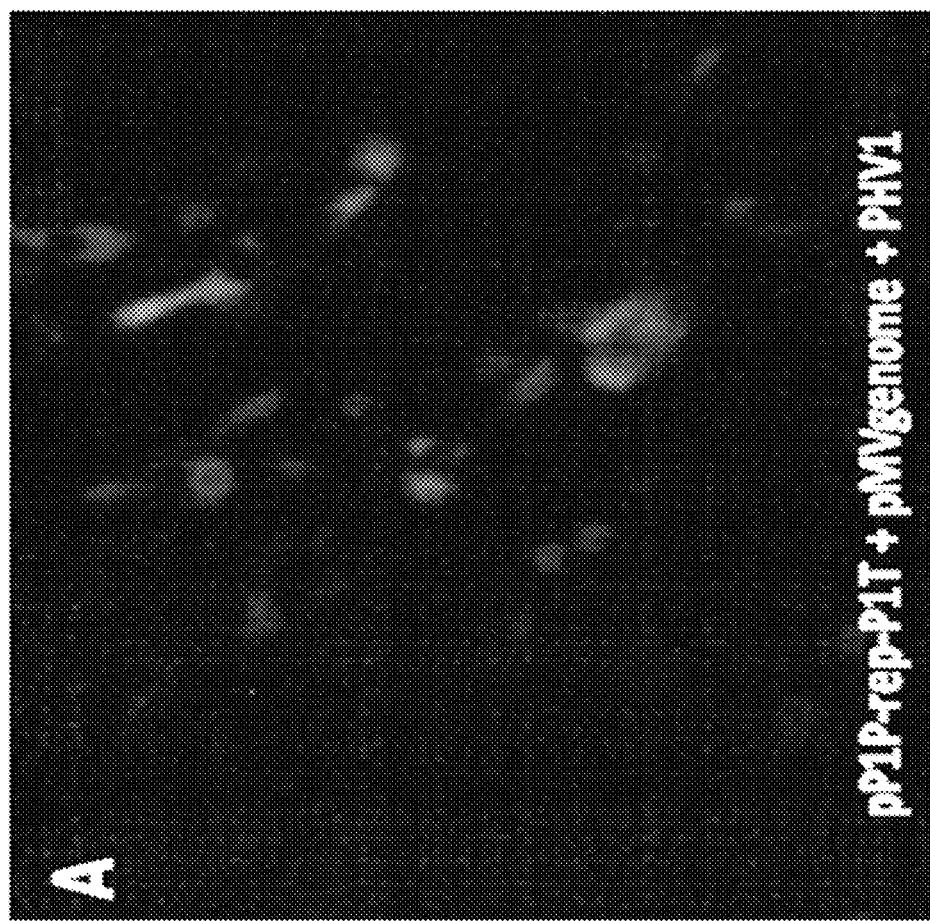


Fig 6A

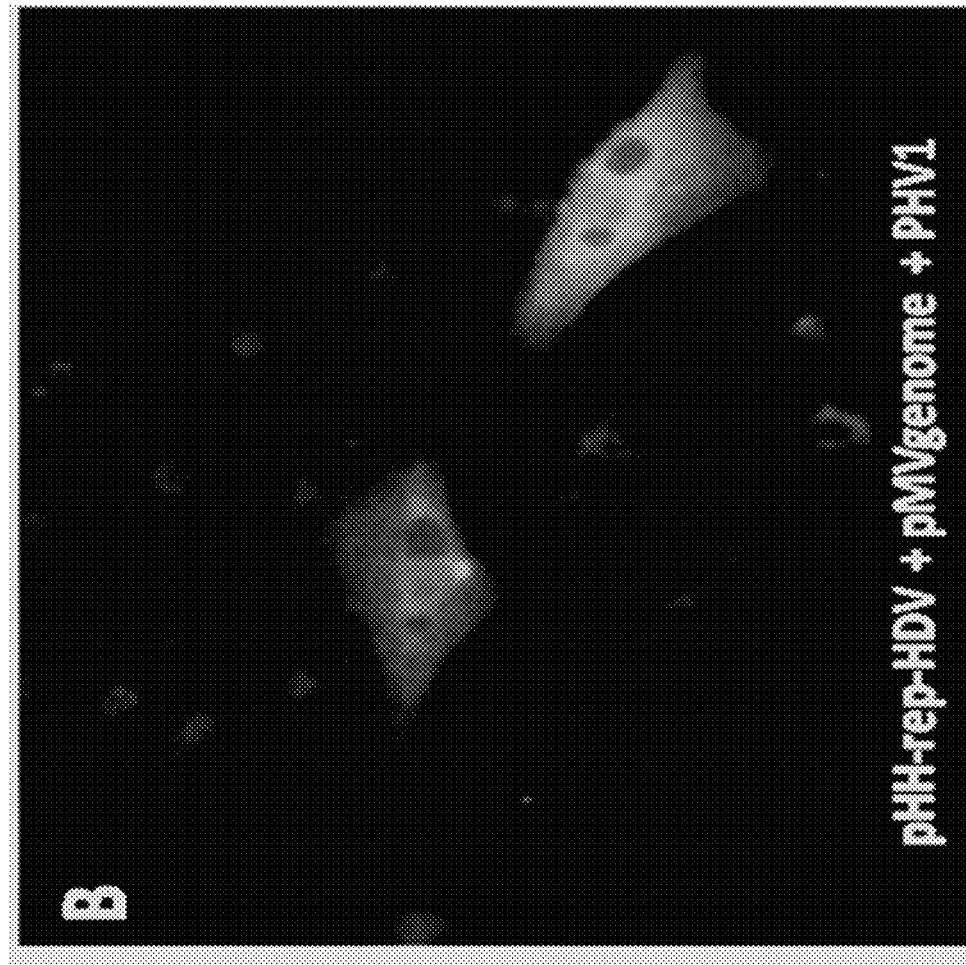


Fig 6B

TWO PLASMID MAMMALIAN EXPRESSION SYSTEM

This application is the U.S. National Phase Application of PCT/IN2012/000405, filed Jun. 8, 2012, which claims priority to Indian Patent Application No. 1679/MUM/2011, filed Jun. 8, 2011, the contents of such applications being incorporated by reference herein.

FIELD OF INVENTION

The present invention relates to a two plasmid mammalian expression system. Moreover invention relates to production of recombinant proteins and viruses. Moreover the present invention relates to a mammalian expression system incorporating methodology for reconstitution of Ribonucleic acid (RNA) dependent RNA polymerase enzyme of negative stranded RNA viruses and its exploitation as a mammalian expression system for the production of proteins, RNA molecules and recombinant viruses.

BACKGROUND OF INVENTION

Advances in molecular biology and genetic engineering led to the development of "Reverse Genetics", a process of generating recombinant viruses from a cloned complementary DNA (cDNA) copy of a viral genome. It has helped understand the molecular determinants related to virus attenuation, tissue tropism virulence factors and in recent years, accelerated the development of virus vaccines by enabling easy modification of viral genomes through manipulation of its cDNA. Reverse genetics has made it possible to produce recombinant viruses with attenuating mutations or chimeric viruses expressing heterologous genes for use as new viral vaccines or therapeutic agents. Morbilli Viruses (Measles Virus and Rinderpest Virus)

MV and RPV are members of the genus morbillivirus of family Paramyxoviridae. Their genetic information is encoded on a single stranded RNA genome of antisense polarity and comprises 15894(MV) and 15882 (RPV) nucleotides respectively. Their genome has unique highly conserved 3' and 5' termini called leader and trailer respectively and encodes 6 genes—N (nucleocapsid protein), P (phosphoprotein), M (matrix protein), F (fusion protein), H (hemagglutinin) and L (large protein=polymerase)—separated by similarly conserved intergenic sequences. In infected cells, viral RNA dependent RNA polymerase (RDRP) initiates transcription of genomic RNA from its promoter present within the leader sequence and produces messenger ribonucleic acid (mRNA) molecules which are translated into corresponding proteins by the cellular ribosomes. At some point in the viral lifecycle and after adequate pools of viral proteins are synthesized, the RDRP enzyme switches mode and initiates replication from another promoter present within the leader sequence of the genomic RNA. Although the exact mechanism(s) regulating the initiation of transcription or replication of viral RNA and transcription start and stop at gene boundaries is poorly understood, the conserved sequences which serve as promoters for transcription and replication and sequences which dictate the transcription start and end at each gene boundary are clearly defined in case of most negative stranded RNA viruses in general and MV and RPV in particular. It is well established that addition of these sequences to any unrelated RNA molecule forms a "virus genome like replicon" which can be transcribed and replicated by its cognate RDRP.

MV causes an acute febrile illness in infants and young children. Its prevalence can be controlled very effectively by vaccination. Most of the currently used live attenuated vaccines including the Schwartz, Moraten, and Edmonston-Zagreb strains are derived from the original Edmonston strain (Enders and Peebles, 1954) by multiple passages in non human cells (Enders, 1962). However, according to the estimates of the World Health Organisation (WHO), one million young children die every year from measles mainly in developing countries. But in recent years developed countries such as the USA with incomplete adherence to vaccination have seen emergence of measles related deaths (Clements and Cutts, 1995). For a recent discussion of MV vaccinology including future trends see Norrby (1995).

During the past 60 odd years, Measles vaccine has been administered in over 700 million children and has proved to be highly effective, usually providing life-long immunity against MV reinfection.

RPV causes cattle plague—an infectious viral disease of cattle, buffalo and other wild life species and is mainly India, Africa and other tropical countries. It is characterized by fever, oral erosions, diarrhea, lymphoid necrosis and high mortality. Two vaccines—Plowright (Plowright and Ferris, 1962) and Lapinized (Scot 1963) have been widely used to protect against rinderpest. The Plowright vaccine derived by attenuation of RBOK strain of RPV has proved to be most effective (Baron et al, 2005). Wide spread use of these vaccines helped irradicate RPV from several countries including India by 2000. However, its reemergence in 2003 has led to resumption of mass vaccinations of cattle and other susceptible animals (Kock et al, 2006).

The proven safety and efficacy of these vaccines, supports their use as an ideal vector for the expression of heterologous genes. Reverse genetics offers a powerful approach for developing recombinant MV or RPV useful as potential vaccines against unrelated diseases and/or therapeutic agents in man and animals.

Reverse genetics was first used to generate RNA viruses by Racaniello and Baltimore in 1981 in case of Poliovirus. Subsequently, several other positive-sense RNA viruses were generated using synthetic RNA produced by T7 or T3 RNA polymerase (Racaniello, V. R. & Baltimore, D., 1981) from a cloned cDNA. Generation of negative stranded RNA viruses however, proved more difficult. Unlike positive stranded RNA viruses, the genome of negative sense viruses cannot be translated by host cells and is not infectious. It must be supplied in the form of ribonucleoprotein (RNP) complexes containing the nucleoprotein and the viral RDRP proteins to allow its transcription and replication and subsequent virus formation. Enami et al, (1990) developed the first reverse genetics system to produce influenza virus (which consists of 9 genomic RNA subunits). Its RNPs are small in size and can be assembled in vitro from RNA and required viral proteins—N and the polymerase components. Initially an artificial RNA carrying a reporter gene—chloramphenicol acetyl transferase (CAT) sequence embedded in viral non-coding terminal sequences of the influenza virus genome subunit was used (Luytjes et al., 1989). Later, single authentic or altered genome subunit RNAs transcribed in vitro from cloned DNA were also used (Enami and Palese, 1991). The assembled RNPs replicated and transcribed upon transfection into influenza-infected cells, as monitored by CAT production and by rescue of a influenza virus, respectively. Purification of virus containing the introduced subunit from the vast excess of non-reassorted virus in some cases can be accomplished by selection, for

example, using a specific neutralising antibody directed against the protein encoded by the cognate subunit of the helper virus.

The RNP's of nonsegmented negative-strand RNA viruses (Mononegavirales) contains in addition to N protein, the assembly and polymerase cofactor phosphoprotein (P) and the viral RNA polymerase (large protein L) and are more difficult to assemble in vitro from synthetic RNA and individual proteins. Therefore, many researchers preferred to use smaller subgenomic RNAs (viral minigenomes) containing the essential sequences of viral genome produced during virus lifecycle were used. They were then substituted by artificially transcribed RNA molecules from DNA constructs containing reporter genes and viral essential non-coding sequences (replicons). Replication of such replicons carrying the CAT coding sequence and viral noncoding terminal sequences was achieved for Sendai virus (Park et al., 1991), Sendai virus (SeV), respiratory syncytial virus (Collins et al., 1993; Collins et al., 1991), human parainfluenza virus 3 (Dimock and Collins, 1993), rabies virus (RV) (Conzelmann and Schnell, 1994) and MV (Sidhu et al., 1995).

A similar system was used to rescue vesicular stomatitis virus (VSV) (Lawson et al., 1995; Schnell, et al, 1994) and rabies virus (RV) entirely from a full length cDNA clone of viral genome under the control T7 RNA polymerase promoter. The components of the viral polymerase complex including the nucleoprotein (NP) were provided from protein expression plasmids that were controlled by T7 RNA polymerase promoter. Soon other researchers also reported generation of non-segmented negative-sense RNA viruses from cloned genomic cDNA for vesicular stomatitis virus (Whelan, et al, 1995), measles virus (Radecke et al, 1995), respiratory syncytial virus (Collins, et al, 1995), sendai virus (Garcin, et al, 1995; Kato et al, 1996), rinderpest virus (Baron & Barrett 1997), human parainfluenza virus (Hoffman et al, 1997; Durbin et al, 1997), simian virus (He et al, 1997), newcastle disease virus (Peeters, et al, 1999) and human severe acute respiratory syndrome corona virus (Yount, et al, 2003).

These demonstrations and other studies of reconstitution of RNA dependent RNA polymerase (RDRP) enzyme activity and its ability to rescue corresponding RNA viruses or non-viral reporter proteins from minireplicons have establish the RDRP enzyme as a powerful versatile system for expression of recombinant proteins either alone or as integral parts of rescued viruses.

The most common methodology used for this purpose, uses transfection of multiple plasmids—one expressing the substrate RNA (a cDNA encoding viral genome or an artificial replicon) and others expressing the viral RDRP complex proteins—viz. the nucleocapsid (N or NP protein), the phosphoprotein (P) and the large polymerase (L) protein and an external T7 RNA polymerase (T7RNAP) to allow expression from these plasmids. The T7 RNAP is used for multiple reasons—(1) its high efficiency, (2) its ability to synthesize RNA with correct 5' terminus identical to viral genome and (3) its ability to transcribe DNA molecules within the cytoplasm thus eliminating modifications of vRNA by RNA splicing, polyadenylation or other mechanisms.

T7RNAP is not a mammalian enzyme. Therefore Pattnaik et al, (1990) used a recombinant attenuated vaccinia virus (VV) (e.g. MVA/T7). It was used for recovery of VSV (Lawson et al, 1995) and rabies virus (Conzelman, U.S. Pat. No. 6,033,886), RSV (Collins et al, 1995); the SV5 (He et al, 1997), HPIV-3 (Durbin et al, 1997), rinderpest virus

(Barn and Barrett 1997) and measles virus (Schneider et al, 1997), mumps virus (Clarke et al, 2000), CDV (Gassen et al, 2000), HPIV-2 (Kawano et al, 2001) and BPIV-3 (Schmidt et al, 2000). Similarly, a recombinant fowlpox virus expressing T7RNAP has also been used to supply T7RNAP for recovery of newcastle virus (NDV) (Peeters et at. 1999) and of a chimeric rinderpest virus (Das et al. 2000).

The recombinant viruses produced using this approach are mixed with vaccinia virus and are difficult to purify which can be a major problem—especially if the recombinant viruses are required for preparing immunogenic compositions or gene therapy vectors. Moreover, this helper vaccinia virus kills the host cells limiting the efficiency of recombinant virus production. Therefore, it would be desirable to eliminate the use of helper virus supplying T7 RNA polymerase. Three different approaches have been used to eliminate the use of externally supplied T7RNAP altogether.

Radecke et al, (1995) produced a helper cell line constitutively expressing T7RNAP and Measles virus (MV) N and P proteins (WO 97/06270) and introduction of a plasmid encoding the entire (+) strand sequence of MV genome linked to T7RNAP promoter and another plasmid encoding MV L protein alone is sufficient to rescue recombinant MV. However, the efficiency of this helper cell line is usually limited and requires to be enhanced by giving a heat shock (Parks et al, 1999). Also, this cell line is only useful for rescue of MV. In contrast, the helper BHK-21 cell line (BSR 17/5) stably expresses only the T7RNAP and can be used for rescue of different viruses as shown in case of BRSV (Buchholz et al. 2000), rabies virus (Finke and Conzelmann 1999), VSV (Harty et al. 2001), NDV (Romer-Oberdorfer et al. 1999), and Ebola virus (Volchkov et al. 2001). It can be used to reconstitute RDRP of any virus by co-transfected with plasmids encoding appropriate N, P and L proteins.

Second approach involves the use of RNA polymerase I (RNAPI). RNAPI is usually involved in transcription of ribosomal genes in mammalian cells. The RNAs synthesized by RNAPI do not contain the 5' methyl cap structure and 3' poly-A tail. The transcription initiation and termination signals for RNAPI are precisely defined and RNA molecules produced by inserting viral genomic or genome like cDNA molecules in between rRNA promoter and terminator signals possess authentic viral 5' and 3' ends, does not require further processing and can be used as a substrate directly by viral RDRP if expressed. (Zobel et al, 1993, Nucleic acids research, 21:3607-3612; Flick and Petterson, 2001, J. Virol. 75: 1643-1655.). Therefore, RNAPI transcription has been used to synthesize viral genomic or genome like cDNA from plasmids and used for rescue of viruses in case of Influenza virus (Neumann et al, 1999), Borna disease virus and MV (Martin et al, 2006,1 Viral. 80:5708-5715).

More recently, Martin et al, (2006) have used a third strategy to express viral genomic RNA from transcripts produced by RNA polymerase II (RNAP II). They placed a hammerhead ribozyme immediately upstream of and a genomic hepatitis delta virus ribozyme immediately downstream of the virus genomic sequence. These ribozymes cleaved a genomic RNA with authentic 3' and 5' ends from the RNA transcribed by RNAP II.

Such strategies eliminate the need for helper virus but still require separate helper plasmids expressing the viral N, P and L proteins. Transfection of so many plasmids simultaneously in a cell and ensuring useful levels of expression of the desired proteins for efficient reconstitution of RDRP can be difficult. Availability of a single helper plasmid expressing all desired genes will help increase the efficiency of virus rescue by ensuring that all transfected cells will receive the

entire complement of helper proteins necessary for reconstitution of RDRP enzyme activity.

This requirement for multiple plasmids has also restricted the use of RDRP based systems to virus rescue, where as studies with artificial replicons encoding reporter proteins has shown that RDRP mediated expression systems can allow high levels expression of recombinant proteins. Availability of a single helper plasmid/reagent to supply the required N, P and L proteins will help expand the scope of using RDRP enzyme for large scale expression of recombinant proteins. Therefore, there exists a need in the art for new simpler methods and reagents which will allow efficient reconstitution of RDRP activity and its exploitation for expression of recombinant proteins, RNA molecules and/or rescue of recombinant viruses.

Here, we describe the preparation and use of simple easily manipulatable plasmid vector systems which can be used for reconstitution of RDRP enzyme activity and its rescue for expression of recombinant proteins, RNA molecules and rescue of recombinant viruses. For this purpose, we have used the RDRP system of 2 viruses—Measles virus (MV) and Rinderpest virus (RPV) as models. These plasmids can be easily modified to express either non-viral proteins, RNA molecules or the entire viral genomes. This vector system will be useful in development of applications related to protein expression and/or generation of recombinant modified viruses (virus rescue) expressing additional proteins and/or RNA molecules useful for vaccination or other therapeutic purposes.

4. OBJECT OF THE INVENTION

The main object of the present invention is to provide two plasmid mammalian expression system for production of recombinant proteins and viruses

Another object is to provide a method for reconstitution of RNA dependent RNA polymerase and its exploitation as a mammalian expression system.

A further object of the inventions is to provide a mammalian expression system for the expression of recombinant proteins, nucleic acid, viruses, RNA molecules.

Still further object of the invention is to provide a mammalian expression system for the intracellular expression of RNA molecules like aptamers, antisense RNA, miRNA, siRNA, ribozymes etc

Yet another object of the invention is to provide reagents for production of recombinant viruses useful as vaccines or therapeutic agents.

Another object of the invention is to describe the process of the preparation of such a mammalian expression system.

5. SUMMARY

The present invention features the use of RNA dependent RNA polymerase enzyme of morbilliviruses for expression of proteins, RNA molecules and production of recombinant viruses in mammalian cells. In one aspect this provides a plasmid DNA molecules which expresse the N, P and L proteins of MV. In another aspect of this invention, it provides another plasmid which expresses easily manipulatable RNA substrate of RDRP which can be used for production of any protein, RNA or modified virus. These plasmids may be used as a reagent kit for expression of proteins or RNA molecules or production of recombinant viruses or combination thereof. Further this invention provides a method for using these plasmids for intracellular expression of RNA molecules which may be useful for

modulation of cellular gene expression. These plasmids may be used in the form of cloning kits.

The following terms/abbreviations used in the specification have meanings attributed to them as mentioned hereinbelow.

MV: Measles Virus, RPV: Rinder Pest Virus, RNA: Ribonucleic acid, DNA: Deoxyribonucleic acid, RDRP: RNA Dependent RNA Polymerase, cDNA: Complimentary DNA, -VRNA: Negative sense Viral RNA, RNP: Ribonucleoprotein, P: Phospho Proteins, L: Large Polymerase Proteins, N: Nucleocapsid, CMV:Cytomegalovirus, IRES: Internal Ribosomal Entry Site, CHO Cell Line: Chinese Hamster Ovary Cell Line, RNA Pol I- RNA Polymerase I, MOI: Multiplicity of Infection, siRNA: selective interfering RNA, miRNA: micro RNA, GFP: green fluorescent protein, HGH: human growth hormone

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1: Schematic diagram of the cloning plasmids encoding the Measles Minireplicon encoding 2 reporter genes.

A. Basic replicon Design

B. Construct 1: HH-replicon-HDV

C. Construct 2: P1P-replicon-P1T

FIG. 2: Schematic representation of the Cloning Plasmids created.

Cloning Plasmid 1: Vector pUC57 was used to clone P1P_RepliCon_P1T construct.

Cloning plasmid 2: Vector pIRES was used to prepare pIRES_P1P_RepliCon_P1T construct.

Cloning plasmid 3: Vector pIRES was used to prepare pIRES_HH_RepliCon_HDV.

FIG. 3: Synthesis of cDNA of entire MV-E genome

FIG. 3a: Generation of cDNA encoding the entire antigenome of MV-E: Viral RNA was purified using Genejet RNA purification kit (Fermentas) and reverse transcribed using Superscript II and random hexamer primers. This was used to amplify seven overlapping fragments with Superscript III and specific primers and cloned into pCDNA3.1 in which the multiple cloning site was replaced with a Nhe I_Not I_Pac I_Pme I linker.

FIG. 3b: Plasmid encoding cDNA of MV-E genome: cDNA encoding entire antigenome of MV-E was synthesized by assembling seven overlapping PCR amplified fragments and cloned in Not I and Pme I sites of pCDNA 3.1(-)

FIG. 4: Schematic representation of the two variants of Helper plasmid created.

Helper Plasmid 1: Vector pBiCMV-1 was used to prepare piCMV_MV-N_MV-P_IRES_MV-L.

Helper plasmid 2: Vector pIRES was used to prepare pIRES_MV-N_p2A_MV-P_MV-L.

FIG. 5: Vero cells were co-transfected with Cloning plasmid encoding eGFP and HGH and HPV1 or HPV 2, incubated for 48 hrs at 37° C. and observed for fluorescence and HGH. A: pUC 18 alone; B: pGFP (positive control); C: pUC_P1P-repliCon-P1T alone; D: pUC-P1P-repliCon-P1T and Helper HPV; E: pIRES-HH-repliCon-HDV alone; F: pIRES-HH-repliCon-HDV and HPV. Note: Helper plasmids 1 and 2 both were able to supply the N, P and L proteins. Representative pictures of HPV1 alone are shown as they were similar.

FIG. 6: Rescue of segmented MV: Equal quantities of plasmids pCDNA_MVgenome, Cloning plasmid 1 (pIRES_HH-repliCon-HDV) and HPV 1 were cotransfected

into Vero cells using Xfect, incubated at 37° C. and observed daily for formation of syncytia. MV-E was harvested from the culture supernatant after syncytia formation covered >80% -90% and titrated using TCID50. Cells were observed simultaneously for expression of EGFP plasmid.

- A: Vero cells transfected with pUC-P1P-rep-P1T, pCDNA-MVgenome & Helper plasmid variant 1;
- B: Vero cells transfected with pIRES-HH-rep-HDV, pCDNA-MVgenome & Helper plasmid variant 1.

DETAILED DESCRIPTION AND EXAMPLES

The present invention relates to expression system that can be easily used for cellular reconstruction of the RDRP enzyme activity for the expression of recombinant proteins or virus rescue. It comprises of two plasmids-1. helper plasmid which expresses N, P and L proteins of MV virus and 2. a Cloning Plasmid which expresses easily manipulatable viral RNA or viral like RNA molecule (minireplicon). The cloning plasmid contains multiple cloning sites (MCS) for easy insertion of DNA encoding target molecule to be expressed. Here MV and RPV are used as model system.

EXAMPLES

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

1. Cells and Viruses

Vero (African green monkey kidney) cells were grown as monolayers in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum (FCS). MRCS cells were grown as monolayers in DMEM supplemented with 10% FCS. Measles virus (Edmonston) (MV-E) strain was purchased from Serum Institute of India (MVAC, 10³ TCID50/vial). To prepare a seed stock, Vero or MRCS cells were seeded in 25 sq. cm flasks at 10⁵ cells/flask and incubated for 36 hrs. Cells were then washed with HBSS and seeded with MV-E at a MOI of 0.1 and supplemented with serum free DMEM. Virus was harvested at 24 hr intervals. Virus collected over 72 hrs was pooled together, quantitated and used as seed stock.

2. Plasmid Constructions

2.1 Cloning Plasmid

2.1.1 Designing the replicon construct

The MV leader (ntds 1 to 107), MV trailer (ntds 15786 to 15894) and the intergenic region between the protein coding regions for MV-N and MV-P proteins (ntd. No 1686 to 1806) were selected from the AY 486084.1 sequence from Genbank (Baricevic et al, 2005). Coding regions for the green fluorescence protein (eGFP) and human growth hormone

(HGH) to be used as reporter proteins were isolated from U55762.1 and NM-000515.3 respectively. All these sequences were assembled in silico into a MV-E genome like replicon containing 2 gene cassettes. Nucleotide sequences corresponding to recognition sites for Afe I, Age I, Asc I, Mlu I, Nru I, Pci I, Sac II, Xho I, Eco RI, Pac I, Pme I, Pml I, Sbf I and Xba I were arranged into 2 oligonucleotides to synthesize 2 multiple cloning sites (MCS1 and MCS 2) and inserted in the replicon around the EGFP and HGH genes. As a result, the EGFP protein appeared to have been cloned within the MCS1 region at Asc I site and HGH protein within MCS 2 at Pac I site (FIG. 1A). The sequence of the replicon without the reporter genes is given in Seq ID No. 1.

The sequence corresponding to a 5' hammer head ribozyme was reconstructed from Combret et al, (2003) and attached at the 5' end of the replicon. Similarly, sequence for 3' hepatitis delta virus ribozyme was adopted from Walker et al, (2003) and appended at the 3' end of replicon to generate a HH-replicon-HDV construct (FIG. 1B; Seq ID No. 2)

Sequences encoding the promoter for chinese hamster RNA polymerase I (P1P) was selected from Tower et al, (1989) and a terminator sequence for murine RNA polymerase I terminator (P1T) was selected on basis sequences described by Grummt et al, (1985, 1986). The P1P sequence was added at immediately upstream of the 5' terminus (immediately upstream) and P1T sequence was appended immediately downstream of the 3' terminus of replicon to create P1P-replicon-P1T construct (FIG. 1C; Seq ID No. 3). 2.1.2 Synthesis of Cloning Plasmids Sequences corresponding to the HH-replicon-HDV (between Eco RI & Hind III sites) and P1P-replicon-P1T (between Sac I and Hind III sites) were synthesized using the gene synthesis method of Young and Dong (2004) and cloned into pUC57 to create pUC_HH-replicon-HDV and pUC_P1P-replicon-P1T respectively.

They were then subcloned in between Nhe I and Not I sites of pIRES vector from Clonetech to generate pIRES_HH-replicon-HDV and pIRES_P1P-replicon-P1T plasmids. These plasmids were used for testing the RNA dependent RNA polymerase (RDRP) mediated expression of GFP and HGH proteins in mammalian cells.

After confirming that these plasmids expressed GFP and HGH under the control of RDRP, the genes for EGFP and HGH were removed by sequential digestion and ligation with Asc I and Pac I to create 3 variants of, cloning plasmids—Cloning Plasmid variant 1 (HH-replicon-HDV) and Cloning plasmid variant 2 (P1P-replicon-P1T) and Cloning Plasmid Variant 3 (pUC_P1P-replicon-P1T). The different plasmids created are listed in Table 1.

TABLE 1

| Different Minireplicon plasmids created. | | | |
|--|---------------------------------------|---|--------------|
| No | Name | Description | Sequence No |
| 1 | Cloning plasmid 1 (pUC-P1P-Rep-P1T) | Replicon under the control of CHO cellular RNA polymerase I promoter and murine RNA polymerase I terminator in pUC57 without reporter genes | Seq ID No. 6 |
| 2 | pUC-HH-Rep-HDV | Replicon flanked by Hammerhead and Hepatitis Delta virus ribozymes at the 5' and 3' termini and cloned in pUC57 vector without reporter genes | Seq ID No. 7 |
| 3 | Cloning plasmid 2 (pIRES-P1P-Rep-P1T) | Replicon under the control of CHO cellular RNA polymerase I promoter and murine RNA polymerase I terminator subcloned into the Nhe I | Seq ID No. 5 |

TABLE 1-continued

| Different Minireplicon plasmids created. | | | |
|--|---|---|--------------|
| No | Name | Description | Sequence No |
| 4 | Cloning plasmid 3 (pIRES-HH-Rep-HDV) | and Not I sites of pIRES vector from Clonetech without reporter genes Replicon flanked by Hammerhead and Hepatitis Delta virus ribozymes at the 5' and 3' termini subcloned into the Nhe I and Not I sites of pIRES vector from Clonetech without reporter genes | Seq ID No. 4 |
| 5 | Cloning Plasmid 1 with reporter genes | Replicon containing reporter genes eGFP and HGH under the control of CHO cellular RNA Polymerase I promoter and murine RNA polymerase I terminator cloned in pUC57 | |
| 6 | Cloning Plasmid 2 with reporter genes | Replicon containing reporter genes eGFP and HGH under the control of CHO cellular RNA Polymerase I promoter and murine RNA polymerase I terminator subcloned into Nhe I and Not I sites of pIRES vector from Clonetech | Seq No F |
| 7 | Cloning Plasmid 3 with reporter genes | Replicon containing reporter genes eGFP and HGH flanked by Hammerhead and Hepatitis Delta virus ribozymes at the 5' and 3' termini subcloned into the Nhe I and Not I sites of pIRES vector from Clonetech | Seq No. G |

2.1.3 Synthesis of cDNA of entire MV-E genome

The MV-E cDNA was cloned from viral particles purified from a batch of MV-E vaccine purchased from the Serum Institute of India, Pune, India. Viral RNA was extracted from 10.sup.5 lysed virus particles using GeneJet RNA purification kit (Fermentas) according to the manufacturer's RNA purification kit according to the manufacturer's protocol. The viral RNA was reverse transcribed into cDNA using random hexamers and Superscript II DNA polymerase. As Seven overlapping cDNA fragments covering the entire viral genome (as shown in FIG. 3a) were generated by PCR using PfuTurbo DNA polymerase and the following primers

- (SEQ ID NO: 10)
(1) 5'-GCGGCCGCACCAAAAC-3';
- (SEQ ID NO: 11)
(2) 5'-CCTGACCGCGGATGC-3';
- (SEQ ID NO: 12)
(3) 5'-ACCTCGCATCCGGCGG-3';
- (SEQ ID NO: 13)
(4) 5'-CCTCCAGAGTAATCGATTAAAGG-3';
- (SEQ ID NO: 14)
(5) 5'-AATCGATTACTCTGGAGGAGCAG-3';
- (SEQ ID NO: 15)
(6) 5'-CTTGCACCTAACITTTAACTAC-3';
- (SEQ ID NO: 16)
(7) 5'-GAACAATATCGGTAGTTAAATTAACAC-3';
- (SEQ ID NO: 17)
(8) 5'-TGAGGGACTCGAGCATACTC-3';
- (SEQ ID NO: 18)
(9) 5'-ATAAGATAGTAGCCATCCTGGAGTAT-3';
- (SEQ ID NO: 19)
(10) 5'-GTAGGGCCATGTGCTGGG-3';
- (SEQ ID NO: 20)
(11) 5'-CATAGCCGTAAACAAAAAGGGTAC-3';
- (SEQ ID NO: 21)
(12) 5'-GAGCATCAAGTGAAGGACCATG-3';

-continued

(SEQ ID NO: 22)

(13) 5'-GCATTGIGGTATTATAGAGCCTATC-3';

(SEQ ID NO: 23)

(14) 5'-CGGTTAAACCAGACAAAGCTG-3'

30 The multiple cloning site from the plasmid pCDNA3.1(-) was removed by digestion with Nhe I and Pme I and replaced it with a linker containing Nhe I-Not I-Pac I-Pme I sites pCDNA-Not_Pac_Pme. The fragments generated by using different primer pairs 7, 8 (Pac I, Xho I), 9,10 (Xho I, Kpn I), 11,12 (Kpn I, Nco I) and 13, 14 (Neo I, Pme I) and ligated into a Pac I-Pme I digested pCDNA-Not_Pac_Pme to generate a plasmid with the nucleotides from Pac I to the 3' end of the MV-E antigenome called pCDNA_Non_Pac MVg_Pme. The fragments generated by other three pairs- 1,2 (Not I, Sac II), 3, 4 (Sac II, Cla I), 5, 6(Cla I, Pac I) were ligated into the Not I-Pac I digested pCDNA_Non_PacM-Vg_Pme plasmid to create pCDNA_MVgenome (FIG. 3b).

2.2 Helper Plasmid

RNA was prepared from the purified MV-E virus purchased from Serum Institute of India, Pune, India using the GeneJet RNA purification kit (Fermentas) according to the manufacturer's protocol. 1.mu.g RNA was reverse transcribed using random hexamers and amplified using primers specific for the N (F: 5'-GCTAGCATGGCCACACttt-AAGG-3' (SEQ ID NO:24) and R 5'-GCGGCCGC-CTAGTCTAGAAGAIT-3' (SEQ ID NO:25)), P (F 5'-GCTAGCATGGCAGAAGAGCAGG-3' (SEQ ID NO:26), R 5'-GCGGCCGCCTACTTCATTATTATC-3' (SEQ ID NO:27)) and L (F 5'-GCTAGCATGGACTCGC-TATCTGTCAAC-3' (SEQ ID NO:28), R 5'-GCGGCCGCT-TAGTCCTTAATCAG -3 (SEQ ID NO:29)) protein coding regions using Superscript III (Invitrogen) as described by Martin et al. (2006) and Combredet et al (2003) using standard molecular cloning techniques. Amplified cDNAs were cloned in between the Nhe I and Not I sites of pIRES vector (Clonetech) to generate pIRES_N, pIRES_P and pIRES_L plasmids.

2.2.1 Synthesis of Helper Plasmid Variant 1

N protein gene was amplified from pIRES_N and subcloned into Eco RI and Pst I sites of pBiCMV1 to generate pBiCMV_N plasmid. The P protein sequence was then

amplified and cloned in Nhe I and Eag I sites to create the pBiCMV_NP construct. The L protein sequence was then subcloned in the Eag I and Sal I sites of pBiCMV_NP plasmid to generate pBiCMV_NPL plasmid. This plasmid contains a bidirectional CMV promoter and can express the N and P proteins. However, the L sequence will be transcribed as a bicistronic RNA with P and will not be translated. Therefore, a mammalian beta globin IRES element (ires) described first, by Chappell et al, (2000) and later on confirmed by Touzlet et al, (2008) to promote efficient translation was inserted immediately upstream of L coding region. An oligonucleotide encoding a pentameric IRES element flanked by a site for Eag I at 5' end and the first 10 nucleotides of L protein at 3' end (5' GGCGTTCTG ACATCCGGCG GTTTCTGAC ATCCGGCGGG 15 TTTCTGACAT CCGCGGGGTT TCTGACATCC GGCGGGGTTTG TGACATCCGG CGGGTGACTC ACAACGGATC CAACAGACAT ATGGACTCGC 3') (SEQ ID NO:30) was synthesized and inserted by site directed mutagenesis into pBiCMV_NPL to generate create pBiCMV_NPiresL plasmid which will also be called Helper Plasmid Variant 1(HPV1) (Seq ID No. 8).

2.2.2 Synthesis of Helper Plasmid Variant 2

N protein sequence was amplified and subcloned in 25 between the Nhe I and Xho I sites to obtain pIRES_N. P protein sequence was then amplified from pIRES_P and cloned into the Eco RI and Mlu I sites to create pIRES_NP. Finally, the L sequence was amplified from pIRES_L and cloned into pIRES_NP between the Sal I and Not I sites to obtain pIRES_NPL. In this form, this plasmid will express N and I proteins but not P. Therefore, a strategy based on the recently described 2A peptide vectors was used to promote the expression of P protein (szymczak and Vignal (2005)). The N and P open reading frames from pIRES_NPL were fused by inserting the oligonucleotide (5' ATCTTCTAGA CGGCTCCGGA GCCACGAAC TCTCTCTGTT AAAGCAAGCA GGAGACGTGG AAGAAAACCC CGGTCCCCATG GCAGAAAGAGC A 3') (SEQ ID NO:31) which encodes the porcine teschovirus 2Apeptide described by Szymczak et al (2007) flanked on the 5' end by the codons immediately before stop codon of MV N protein and on the 3' end by the first few codons of MV P protein by site directed mutagenesis to fuse the N and P protein regions into a single N2AP fusion protein and obtain pIRES_N2aPL plasmid which will also be called Helper Plasmid Variant 2(HPV2) (Seq ID No. 9).

The plasmids HPV1 (pBiCMV_NPiresL) and HPV2 (pIRES_N2aPL) are represented schematically in FIG. 4.

2.2.3 Synthesis of Equivalent Helper Plasmids Encoding N, P and L Proteins of Other Negative Stranded RNA Viruses

The cloning strategy used for generation of these helper plasmids was then tested for its applicability to other negative stranded RNA viruses—mainly MV, Rinderpest (RPV); peste des petits ruminants (PPRV) canine distemper (CDV), newcastle disease (NDV) and sendai viruses (SeV). Coding regions of the nucleocapsid, phosphoprotein and large proteins were analysed for the presence of restriction enzymes Eco RI, Pst I, Nhe I, Eag I, Sal I, Xho I, Mlu I and Not I.

Sites for Eco RI and Pst I were absent in the nucleocapsid proteins of MV and CDV. Similarly, sites for Nhe I and Xho I were absent in the nucleocapsid proteins of MV and SeV. However, variable number of sites for enzymes Eco RI, Pst I, Nhe I and Xho I were detected in the nucleocapsid of other viruses (Table 2).

TABLE 2

| Presence of sites for Eco RI, Pst I, Nhe I and Xho I in the N protein of various negative stranded RNA viruses | | | | | |
|--|-------------|--------|-------|-------|-------|
| Virus | Gen bank No | Eco RI | Pst I | Nhe I | Xho I |
| MV | AY 486084.1 | 0 | 0 | 0 | 0 |
| RPV | AB 547190.1 | 1 | 3 | 1 | 0 |
| PPRV | HQ197753.1 | 0 | 2 | 0 | 1 |
| CDV | AB 687721.2 | 0 | 1 | 2 | 0 |
| NDV | HQ008337.1 | 0 | 0 | 1 | 2 |
| Sendai | NC_001552.1 | 0 | 1 | 0 | 0 |

Sites for enzymes Eag I, Sal I and Not I were absent from the L proteins of MV, PPRV, CDV and NDV. RPV and SeV contained 1 site for Sal I in their L proteins (Table 3).

TABLE 3

| Presence of sites for Eag I, Sal I and Not I in the L proteins of various negative stranded RNA viruses | | | | |
|---|-------------|-------|-------|-------|
| Virus | Gen bank No | Eag I | Sal I | Not I |
| MV | AY 486084.1 | 0 | 0 | 0 |
| RPV | AB 547190.1 | 0 | 1 | 1 |
| PPRV | HQ197753.1 | 0 | 0 | 0 |
| CDV | AB 687721.2 | 0 | 0 | 0 |
| NDV | HQ008337.1 | 0 | 0 | 0 |
| Sendai | NC_001552.1 | 0 | 1 | 1 |

However, the genes for both these proteins encode a 30 single protein each. Therefore, it would be easily possible to make synonymous mutations in their protein coding regions and eliminate the sites for these restriction enzymes. Therefore, the same cloning strategy can be easily used to clone the nucleocapsid and Large protein coding regions in a helper plasmid construct similar to either helper plasmid variant 1 or helper plasmid variant 2.

Similar to the above results, analysis of the phosphoprotein coding regions of these viruses revealed the presence of 40 a variable number of sites for enzymes Nhe I, Eag I, Eco RI and Mlu I (Table 4). Sites for these enzymes were absent from the phosphoprotein coding regions of MV, CDV and NDV.

TABLE 3

| Presence of sites for Nhe I, Eag I, Eco RI and Mlu I in the P of various negative stranded RNA viruses | | | | | |
|--|-------------|-------|-------|--------|-------|
| Virus | Gen bank No | Nhe I | Eag I | Eco RI | Mlu I |
| MV | AY 486084.1 | 0 | 0 | 0 | 0 |
| RPV | AB 547190.1 | 1 | 0 | 3 | 0 |
| PPRV | Z30697.2 | 0 | 0 | 0 | 0 |
| CDV | AB 687721.2 | 0 | 0 | 0 | 0 |
| NDV | HQ008337.1 | 0 | 0 | 0 | 0 |
| sendai | NC_001552.1 | 0 | 0 | 1 | 0 |

Although the P protein of RPV (AB547190) sequence is digested by Nhe I and Eco RI the regions corresponding to the recognition sites of these enzymes varies across different 60 strains of RPV (e.g. Z30697.2 in Genbank). On the other hand, the Eco RI site in the P protein of PPRV appears to be highly conserved across most PPRV strains. However, this region of the P protein coding sequence does not overlap with the coding regions of C and V proteins which are also coded by the P gene transcript. Thus, it would be possible to introduce synonymous mutations in the P proteins of RPV and PPRV to enable the use of our proposed strategy for

13

preparing the helper plasmids for MV, CDV, RPV, PPRV and NDV. Therefore, the same restriction enzymes may be used to synthesize helper plasmid constructs equivalent to those described as Helper Plasmid Variant 1 and Helper Plasmid Variant 2 from the nucleocapsid (N or NP), phosphoprotein (P) and large (L) proteins of other negative stranded RNA viruses. Such variants will be useful as helper plasmids for reconstitution of corresponding viral RNA dependent RNA polymerase enzyme and its exploitation for protein or RNA expression and also generation of recombinant viruses as novel vaccines and/or therapeutic agents.

3. Expression of Recombinant Proteins by Plasmid Encoded RDRP

First the capacity of cloning plasmids to express RNA molecules which can serve as substrate for MV RNA dependent RNA polymerase (RDRP) was evaluated using a system similar to the one described by Martin et al, (2006). Briefly, Vero cells were transfected with Cloning plasmid, and individual plasmids expressing the N, P and L proteins of MV-E at a ratio of 1:1:1:0.5 in lipofectamine (Invitrogen) according to the manufacturer's protocol. Cells were incubated at 37° C. in 5% CO₂ for 48 hrs and evaluated for expression of green fluorescent protein (eGFP) by microscopy and fluorescence measurement using microplate reader.

In a subsequent experiment, Vero cells were transfected with equal proportions of one cloning plasmid (pUC-P1P-replicon-P1T or pIRES-HH-replicon-HDV or pIRES-P1P-replicon-P1T) and one helper, plasmid (Helper variant 1 or Helper variant 2) in lipofectamine (Invitrogen) or xfect (Clonetech) and incubated at 37° C. in 5% CO₂ for 48 hrs and evaluated for the expression of green fluorescent protein (eGFP) by microscopy and fluorescence.

4. Rescue of MV-E

The capacity of the helper plasmids to rescue MV-E from cDNA was tested. Plasmid pCDNA_MVgenome was cotransfected with Helper plasmid variant 1 or Helper plasmid variant 2 in Vero cells using Xfect and incubated overnight at 37° C. Transfection medium was replaced by fresh medium and cells were incubated further for two days.

14

When syncytia involved 80% to 90% of cell layer, virus was harvested by scraping infected cells, freeze-thawing of cells and medium and centrifugation to remove cellular debris. Collected virus was titrated using the TCID50 titration method. Briefly, Vero cells were seeded into 96 well plate (7500 cells/well) and infected by serial 1:10 dilutions of virus sample in DMEM containing 5% DCS. After incubation at 37° C. for 7 days, cells were stained with crystal violet and virus dilution that resulted in infection of 50% of test unit was determined. The 50% end point described as tissue culture infectious dose (TCID50) was calculated by the Kaber method. Virus rescued from the pCDNA_MVgenome+Helper plasmid had titers of 10⁶ to 10⁷ TCID50/ml.

5. Rescue of Segmented MV-E Using Plasmid Encoded RDRP

The capacity of the helper plasmids to rescue recombinant segmented MV-E from cDNA was tested. Vero cells were cotransfected with pCDNA_MVgenome, Cloning plasmid encoding eGFP and either HPV 1 or HPV 2 in equal proportions using Xfect and incubated overnight at 37° C. Transfection medium was replaced by fresh medium and continued to incubate with daily observation for syncytia formation. When syncytia involved 80% to 90% of cell layer, virus was harvested by scraping infected cells, freeze-thawing of cells and medium and centrifugation to remove cellular debris. Collected virus containing was titrated using the TCID50 titration method. Briefly, Vero cells were seeded into 96 well plate (7500 cells/well) and infected by serial 1:10 dilutions of virus sample in DMEM containing 5% DCS. After incubation at 37° C. for 7 days, cells were stained with crystal violet and virus dilution that resulted in infection of 50% of test unit was determined. The 50% end point described as tissue culture infectious dose (TCID50) was calculated by the Kaber method. Virus rescued from the originally transfected cells had titers of 10⁶ to 10⁷ TCID50/ml. Cells infected with the virus harvested from the originally transfected vero cells also expressed eGFP indicating successful packaging eGFP encoding minireplicon along with MV-E genome into virions and its transfer to fresh cells.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 31

<210> SEQ ID NO 1
<211> LENGTH: 439
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence assembled from sequences obtained from
different organisms or plasmids

<400> SEQUENCE: 1

accaaacaaa gttgggtaag gatagttcaa tcaatgtca ttttcttagtg cacttaggt 60
tcaagatcct attatcgagg acaagagcag gattaaggat atccgagtcg cgacgcgtac 120
atgttagcgct cgcacccggtc cgcggggcgc gccctcgagg tgcgagaggc cgaggaccag 180
aacaaacatcc gcctaccctc catcatttgtt ataaaaaact taggaaccag gtccacacag 240
ccgcccagccc atcaaccatc cactcccacg attggagccg cacgtgtcta gagggccgt 300
ttaaacccctg caggtaattt aagtgaattc ttgggttgaac tccggaaccc taatctgtcc 360
ctagggttgtt aggcatattt tgcaatagat taaagaaaaac tttgaaaata cgaagttct 420
attccccagct ttgtctgggt 439

```

-continued

<210> SEQ_ID NO 2
<211> LENGTH: 614
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 2

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|-----|
| aagcttgcta | gcaccaactt | tgttgttct | gatgagtccg | tgaggacgaa | accggagtc | 60 |
| ccgggtcacc | aaacaaagtt | gggttaaggat | agttcaatca | atgatcattt | tcttagtgcac | 120 |
| ttaggattca | agatcctatt | atcagggaca | agagcaggat | taaggatatac | cgagtcgcga | 180 |
| cgcgtacatg | tagcgctcgc | accggtccgc | ggggcgcgcgc | ggcgcgcgcct | cgagggtgcga | 240 |
| gaggccgagg | accagaacaa | catccgccta | ccctccatca | ttgttataaaa | aaacttagga | 300 |
| accagggtcca | cacagccgccc | agcccatcaa | ccatccactc | ccacgattgg | agccgcacgt | 360 |
| gtcttagaggg | cccggtttaaa | ccctgcaggt | ttaattaagt | gaatttttgg | ttgaactccg | 420 |
| gaaccctaat | cctgccttag | gtgggttaggc | attatttgca | atagattaaa | aaaaactttg | 480 |
| aaaatacgaa | gtttctattc | ccagctttgt | ctgggtgccgg | ccatggtccc | agccctcccg | 540 |
| ctggcggccgc | gtgggcaaca | ttccgaggggg | accgtccccct | cggtaatggc | gaatgggacc | 600 |
| gcggccgcgca | gctc | | | | | 614 |

<210> SEQ_ID NO 3
<211> LENGTH: 969
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 3

| | | | | | | |
|-------------|------------|-------------|-------------|-------------|------------|-----|
| aagcttgct | agcacatcct | cttggtccta | tcacggttat | gaggtcgacc | agtgttgct | 60 |
| ttgatgttcg | gttctctcg | tgattggac | aatatttggg | gcacttcgccc | ggtcccgact | 120 |
| tccagaattt | ccgtgtggc | tgtgaattt | tcaccgtac | actgtcatca | tattccagtt | 180 |
| ttgcaatctg | ctctctttgt | acctgcagat | aggtacaaaa | caaagttggg | taaggatagt | 240 |
| tcaatcaatg | atcattttct | agtgcactta | ggattcaaga | tcctattatc | agggacaaga | 300 |
| gcaggattaa | ggatatccga | gtcgcgacgc | gtacatgttag | cgctcgcacc | ggtccgcggg | 360 |
| gcgcgcgcctc | gaggtgcgag | aggccgagga | ccagaacaac | atccgcctac | cctccatcat | 420 |
| tgttataaaa | aacttaggaa | ccaggtccac | acagccgcca | gccccatcaac | catccactcc | 480 |
| cacgattgga | gccgcacgtg | tctagagggc | ccgtttaaac | cctgcaggtt | taattaagt | 540 |
| aattcttgg | tgaactccgg | aaccctaattc | ctgcccattag | tggtttaggc | ttatttgcaa | 600 |
| tagattaaag | aaaactttga | aaatacgaag | tttctattcc | cagctttgtc | tggttttttt | 660 |
| cccccccaac | ttcggaggc | gaccagtact | ccgggcgaca | ctttgttttt | tttttttccc | 720 |
| cggatgtgg | aggtcgacca | gatgtccgaa | agtgtcccc | cccccccccc | cccccccccg | 780 |
| gcgcggagcg | gcggggccac | cccgacccc | ttttttttt | ttttttttt | ttttaattc | 840 |
| ctggAACCTT | taggtcgacc | agttgtccgt | cttttactcc | ttcatatagg | tcgaccagta | 900 |
| ctccgggtgg | tactttgtct | ttttctgaaa | atcccagagg | tcgaccagat | atccgcggcc | 960 |
| gcgcgagc | tc | | | | | 969 |

-continued

<210> SEQ ID NO 4
<211> LENGTH: 6023
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 4

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ggccggttcc | cttttagtgag | ggtaatgtc | tcgagcagac | atgataagat | acattgtga | 60 |
| gttggacaa | accadaacta | gaatgcagt | aaaaaaatgc | tttatttgta | aaatttgta | 120 |
| tgcattgtc | ttatgttaa | ccattataag | ctgcaataaa | caagttaca | acaacaattg | 180 |
| cattcattt | atgttcagg | ttcaggggga | gatgtgggag | gtttttaaa | gcaagtaaaa | 240 |
| cctctacaaa | tgtggtaaaa | tccgataagg | atcgatccgg | gctggcgtaa | tagcgaagag | 300 |
| geccgcaccc | atcgcccttc | ccaacagttg | cgcagcctga | atggcgaatg | gacgcccct | 360 |
| gtagcggcgc | attaagcgcg | gccccgtgtgg | tggtaacgcg | cagcgtgacc | gtcacacttg | 420 |
| ccagcgcct | agcgcggcgt | ccttcgcctt | tcttccttc | ctttctcgcc | acgttcgccg | 480 |
| gtttccccc | tcaagctcta | aatgggggc | tccctttagg | gttccgattt | agagettac | 540 |
| ggcacctcga | ccgaaaaaaaa | cttgatttgg | gtgatggttc | acgtatgtgg | ccatcgccct | 600 |
| gatagacggt | tttccgcct | ttgacgttgg | agtccacgtt | cttttaatgt | ggactctgt | 660 |
| tccaaactgg | aacaacactc | aaccctatct | cggtctattt | ttttgatttta | taaggattt | 720 |
| tgccgatttc | ggccttattgg | ttaaaaaatg | agctgatttta | acaaatattt | aacgagaatt | 780 |
| ttaacaaaat | attaacgttt | acaatttcgc | ctgatgegg | attttctct | tacgcatctg | 840 |
| tgccgttattt | cacaccgc | acgcccattt | gcccgcacc | atggcctgaa | ataacctctg | 900 |
| aaagaggaac | ttggtaggt | accttcttag | gcggaaagaa | ccagctgtgg | aatgtgtgc | 960 |
| agtttaggtg | tggaaagtcc | ccaggctccc | cagcaggcag | aagtatgca | agcatgcattc | 1020 |
| tcaattagtc | agcaaccagg | tgtggaaagt | ccccaggctc | cccagcaggc | agaagtatgc | 1080 |
| aaagcatgca | tctcaattag | tcagcaacca | tagccccgc | cctaactccg | cccatccgc | 1140 |
| ccctaactcc | gcccaagtcc | gccattctc | cgcggatgg | ctgactaatt | tttttattt | 1200 |
| atgcagggc | cgaggccgcc | tcggcctcg | agctattcc | gaagtagtga | ggaggcttt | 1260 |
| ttggaggcct | aggctttgc | aaaaagctt | attctctga | cacaacagtc | tgcactaa | 1320 |
| ggctagagcc | accatgattt | aacaagatgg | attgcacgc | ggttctccgg | ccgcttgggt | 1380 |
| ggagaggcta | tccggctatg | actgggcaca | acagacaatc | ggctgtctg | atgcccgcgt | 1440 |
| gttccggctg | ttagcgcagg | ggcccccgg | tcttttgc | aagaccgacc | tgcgggtgc | 1500 |
| cctgaatgaa | ctgcaggacg | aggcagcgc | gctatctgg | ctggccacga | cggcgatcc | 1560 |
| ttgcgcagct | gtgcgcacg | ttgtcaactg | agcggaaagg | gactggctgc | tattggcga | 1620 |
| agtgcgggg | caggatctcc | tgtcatctca | ccttgcctt | gcccggaaag | tatccatcat | 1680 |
| ggctgtatgca | atgcggccgc | tgcatacgct | tgatccggct | acctgccccat | tcgaccacca | 1740 |
| agcgaaacat | ogcatcgagc | gagcacgtac | tcggatggaa | gcccgttctg | tcgatcagg | 1800 |
| tgatctggac | gaagagcatc | aggggctcgc | gccagccgaa | ctgttcgc | ggctcaaggc | 1860 |
| gcgcacatccc | gacggcagg | atctcgctgt | gacccatggc | gatgcctgt | tgccgaatat | 1920 |
| catggtgaa | aatggccgct | tttctggatt | catcgactgt | ggccggctgg | gtgtggcgga | 1980 |
| ccgcatacg | gacatagcgt | tggctacccg | tgatattgt | gaagagcttg | gcggcgaatg | 2040 |

US 9,441,205 B2

19

20

-continued

| | |
|---|------|
| ggctgaccgc ttacgtgc tttacggtat cgccgcgtccc gattcgacgc gcatacgccctt | 2100 |
| ctatcgccctt cttagacgagt ttcttcgac gggactctgg ggttcgaaat gaccgaccaa | 2160 |
| gccccccca acctgcccac acgtggccg caataaaaata tctttatctt cattacatct | 2220 |
| gtgtgttgtt ttttgttg aatcgatgc gataaggatc cgctgtatggt gcactctcag | 2280 |
| tacaatctgc tctgtatccg catagttaag ccagccccga caccggccaa caccggctga | 2340 |
| cgcgcctgaa cggggttgc tgctccggc atccgcttac agacaagctg tgaccgtctc | 2400 |
| cgggagctgc atgtgtcaga gtttttccacc gtcatcacccg aaacgcgcga gacgaaagg | 2460 |
| cctcgtata cgcctatattt tataaggtaa tgtcatataa ataatggtt ctttagacgtc | 2520 |
| aggtggcact ttccgggaa atgtgcgcgg aacccttattt tgtttatttt tctaaataca | 2580 |
| ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa | 2640 |
| aaggaagagt atgagtattc aacatttccg tgctccctt attccctttt ttgcggcatt | 2700 |
| ttgccttcctt gttttgttc acccagaaac gctggtaaa gtaaaagatg ctgaagatca | 2760 |
| gttgggtgcg cagatgggtt acatcgaaact ggatctcaac agcggtaaga tccttgagag | 2820 |
| tttcgcggcc gaagaacgtt ttcaatgtat gagcactttt aaagttctgc tatgtggcgc | 2880 |
| ggtattatcc cgtattgacg cccggcaaga gcaactcggt cgccgcatac actattctca | 2940 |
| gaatgacttg gttgagttt caccagtac agaaaagcat cttacggatg gcatgacagt | 3000 |
| aagagaattt tgcaatgtc ccataaccat gagtgataac actgcggcca acttacttct | 3060 |
| gacaacgatc ggaggaccga aggagctaaccg cgtttttt cacaacatgg gggatcatgt | 3120 |
| aactcgcctt gatcggttggg aaccggagct gaatgaagcc ataccaaaacg acgagcgtga | 3180 |
| caccacgatc cctgtaccaa tggcaacaac gttgcgc当地 ctattaactg gcaactact | 3240 |
| tactcttagt tcccgcaac aattaataga ctggatggag gcccataaag ttgcaggacc | 3300 |
| acttctgcgc tcggcccttc cggctggctg gtttattgtat gataatctg gagccgtga | 3360 |
| gcgtgggtct cgcgttatca ttgcacact gggccagat ggttgcgc当地 cccgtatctg | 3420 |
| agttatctac acgacgggaa gtcaggcaac tatggatgaa cgaaatagac agatcgctga | 3480 |
| gataggtgcc tcactgatta agcattggta actgtcagac caagtttact catataact | 3540 |
| tttagattgt taaaacttc attttaatt taaaaggatc taggtgaaga tccttttga | 3600 |
| taatctcatg accaaaaatcc cttaacgtga gtttgc当地 cactgacgtc cagaccgt | 3660 |
| agaaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtatct gctgtttc | 3720 |
| aacaaaaaaaaa ccacccgtac cagcggtggg ttgtttgc当地 gatcaagagc taccaactct | 3780 |
| ttttccgaag gtaactggct tcagcagac gcagatacca aatactgtcc ttcttagtga | 3840 |
| gcgcgtatgtt ggccaccact tcaagaactc tgcacccg cctacatacc tcgc当地 | 3900 |
| aatccgttta ccaggctggctg ctgc当地 cgttaccg ggttggactc | 3960 |
| aagacgtatc ttacccgata aggccgc当地 gtcgggctga acggggggggg cgtgcacaca | 4020 |
| ccccagctt gaggcaacga cctacaccga actgagatac ctacagcgtg agctatgaga | 4080 |
| aaggccacg cttccgc当地 ggagaaaggc ggacaggatc ccggtaagcg gcagggtcg | 4140 |
| aacaggagag cgcacgagg agcttccagg gggaaacgca tggatcttt atagtcctgt | 4200 |
| cgggttgc当地 cacccgtac ttgc当地 cgttgc当地 gggggccgg | 4260 |
| cctatggaaa aacgecagca acggccctt ttacggctt ctggccctt gctggccctt | 4320 |
| tgctcacatg gctcgacaga tcttcaatat tggccattag ccatattattt catggat | 4380 |
| atagcataaa tcaatattgg ctatggca ttgcatacgt tgc当地 tata tcataatatg | 4440 |

-continued

| | |
|---|------|
| tacattttt ttggcgtatg tccaaatatga ccgcacatgtt ggcattgtt attgactagt | 4500 |
| tattaatagt aatcaattac ggggtcatta gttcatagcc catatatggc gttcccggtt | 4560 |
| acataactta cggtaaatgg cccgcctggc tgaccggcc acgaccccc cccattgacg | 4620 |
| tcaataatga cgtatgttcc catagtaacg ccaataggga cttdccattt acgtcaatgg | 4680 |
| gtggagtatt tacggtaaac tgcccacttg gcagttacatc aagtgttatca tatgccaagt | 4740 |
| ccggcccccta ttgacgtcaa tgacggtaaa tggccgcctt ggcattatgc ccagttacatg | 4800 |
| accttacggg actttcctac ttggcagttac atctacgtat tagtcatcgc tattaccatg | 4860 |
| gtgtatgcgtt ttggcagttac caccaatggg cgtggatagc ggtttactc acggggattt | 4920 |
| ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcacccaaa tcaacgggac | 4980 |
| tttccaaat gtgcgtacaa ctgcgtatcgc ccggcccggtt gacgcacaaatg ggcggtaggc | 5040 |
| gtgtacggtg ggaggcttat ataaggcagag ctgcgtttagt gaaccgtcag atcactagaa | 5100 |
| gttttattgc ggttagtttat cacagttaaa ttgcgtacgc agtcgtgttct tctgcaccaa | 5160 |
| cagtcgtcgttac cttaagctgc agtgcacttc ttaaggtagc ctgcgtacgg ttggcgtgtga | 5220 |
| ggcactgggc aggttaagtat caaggttaca agacaggttt aaggagacca atagaaactg | 5280 |
| ggcttgcgtca gacagagaag actcttgcgtt ttctgtatgg cacctattgg tcttactgac | 5340 |
| atccactttt cttttcttc cacagggttc cactccagt tcaattacag ctcttaaggc | 5400 |
| tagagtactt aatacgactc actataggtt agcaccaact ttgtttggc tggatgttcc | 5460 |
| gtgaggacga aaccggaggtt cccgggtcac caaacaatgtt gggtaagga tagttcaatc | 5520 |
| aatgtatcatt ttctgtgtca cttaggatcc aagatccat tatcaggggac aagagcagga | 5580 |
| ttaaggatattt ccgagtcgc acgcgtacat gtgcgtctcg caccgggtcc cgccgcgcgc | 5640 |
| cgccgcgcgc tcgagggtgcg agaggccgag gaccagaaca acatccgcctt accctccatc | 5700 |
| attgttataa aaaacttagg aaccagggtcc acacagccgc cagcccatca accatccact | 5760 |
| cccacgttgcg gagccgcacg tgcgttaggg gcccgtttaa accctgcagg tttatattgc | 5820 |
| tgaattttt gttgaactcc ggaaccctaa tccgtccatc ggtgggttagg cattatttgc | 5880 |
| aatagattaa agaaaacttt gaaaatacga agtttctattt cccagcttgc tctggcgcgc | 5940 |
| gcacatggtcc cagcccttc gctggcggcc ggtggcAAC attccgaggg gaccgtcccc | 6000 |
| tgcgtatgg cgaatgggac cgc | 6023 |

```
<210> SEQ ID NO 5
<211> LENGTH: 6376
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from
different organisms or plasmids
```

<400> SEQUENCE: 5

| | | | | | | |
|------------|-------------|------------|------------|-------------|-------------|-----|
| ggccgcttcc | cttttagtgag | ggtaatgtc | tcgagcacac | atgataaagat | acatttgatga | 60 |
| gttggacaa | accacaacta | aatgcagtg | aaaaaaatgc | tttatttgtg | aaatttgtga | 120 |
| tgttattgtc | ttatttgtaa | ccattataag | ctgcaataaa | caagtttaca | acaacaattg | 180 |
| cattcatttt | atgtttcagg | ttcaggggaa | gatgtgggag | gtttttaaa | gcaagtaaaa | 240 |
| cctctacaaa | tgtggtaaaa | tccgataagg | atcgatccg | gctggcgtaa | tagcgaagag | 300 |
| gccccacccg | atcgcccttc | ccaacagttg | cgcagcctga | atggcgaatg | gacgcgcct | 360 |
| qtaqcqqcq | attaadqcqq | qcqqqtqfq | tqttacqcq | caqcqtcacc | qctacacttq | 420 |

-continued

| | | | | | | |
|-------------|--------------|--------------|-------------|-------------|-------------|------|
| ccagcgccct | agcgccccgt | cctttcgctt | tcttccttc | ctttctcgcc | acgttcgccg | 480 |
| gtttcccccgt | tcaagctcta | aatccgggggc | tcccttttagg | gttccgattt | agagctttac | 540 |
| ggcacctcga | ccgcaaaaaaa | cttgatttgg | gtgatggttc | acgttagtggg | ccatcgccct | 600 |
| gatagacggt | ttttcgccct | ttgacgttgg | agtccacgtt | ctttaatagt | ggactcttgt | 660 |
| tccaaactgg | aacaacactc | aaccctatct | cggtcttattc | ttttgattta | taagggattt | 720 |
| tgccgatttc | ggcctattgg | ttaaaaaatg | agctgattta | acaaatattt | aacgcgaatt | 780 |
| ttaacaaaat | attaacgtt | acaatttgc | ctgatgcggt | attttctct | tacgcacatcg | 840 |
| tgcggtattt | cacaccgcat | acgccccatct | gegcagcacc | atggcctgaa | ataacctctg | 900 |
| aaagaggaac | ttggtttaggt | accttctgag | gccccaaagaa | ccagctgtgg | aatgtgtgtc | 960 |
| agtttaggggt | tggaaaagtcc | ccagggctccc | cagcaggcag | aagtatgcaa | agcatgcattc | 1020 |
| tcaattagtc | agcaaccagg | tgtggaaagt | ccccaggctcc | cccagcaggc | agaagtatgc | 1080 |
| aaagcatgca | tctcaattag | tcagcaacca | tagtccccgg | ccttaactccg | cccatcccg | 1140 |
| ccttaactcc | gcccaaggttcc | gcccaatttctc | cggccatgg | ctgactaatt | ttttttat | 1200 |
| atgcagaggc | cgaggccgccc | tcggcctctg | agctatttcca | gaagtagtga | ggaggcttt | 1260 |
| ttggaggccct | aggcttttgc | aaaaagcttgc | atttttctga | cacaacagtc | tcgaacttaa | 1320 |
| ggctagagcc | accatgatttgc | aacaagatgg | attgcacgc | ggttctccgg | ccgcttgggt | 1380 |
| ggagaggctt | tccggctatgc | actgggcaca | acagacaatc | ggctgtctg | atgcggccgt | 1440 |
| gttccggctt | tcagegcagg | ggcggccgggt | tctttttgtc | aagacgcacc | tgtccgggtc | 1500 |
| cctgaatgaa | ctgcaggacg | aggcagcgcgc | gctatcgtgg | ctggccacga | cgggegttcc | 1560 |
| ttgcgcagct | gtgcgtcagc | ttgtcactga | agcgggaagg | gactggctgc | tattggcga | 1620 |
| agtgcggggg | caggatctcc | tgtcatctca | ccttgcttcc | gccgagaaag | tatccatcat | 1680 |
| ggctgtatgc | atgcggccgc | tgcatacgt | tgtatccgc | acctgcccatt | tcgaccacca | 1740 |
| agcgaaacat | cgcacatcgac | gagcacgtac | tcggatggaa | gccggctctg | tcgatcgagg | 1800 |
| tgtatctggac | gaagagcatc | aggggctcgc | gccagccgaa | ctgttgcaca | ggctcaaggc | 1860 |
| gcccgcgcgc | gacggcgagg | atctcgctgt | gacccatggc | gatgcctgt | tgccgaatat | 1920 |
| catgggtggaa | aatggccgct | tttctggatt | catcgactgt | ggccggctgg | gtgtggcgga | 1980 |
| ccgctatcag | gacatagcgt | tggctaccgc | tgtatattgc | gaagagctgc | ggggcgaatg | 2040 |
| ggctgaccgc | tccctcgtgc | tttacggtat | ccggcgtccc | gattcgcagc | gcategcctt | 2100 |
| ctatcgccctt | cttgacgagt | tcttctgagc | gggactctgg | ggttcgaaat | gaccgaccaa | 2160 |
| gccccccca | acctgcccattc | acgtggccgc | caataaaata | tctttat | tatttattt | 2220 |
| gtgtgttgtt | tttttgtgtt | aatcgatgc | gataaggatc | cgcgatgtgt | gcactctcag | 2280 |
| tacaatctgc | tctgtatgcgc | catagttaag | ccagccccgc | cacccgcacaa | cacccgcgt | 2340 |
| cgccgcctga | cgggcttgc | tgctccggc | atccgcttac | agacaagctg | tgaccgttcc | 2400 |
| cgggagctgc | atgtgtcaga | ggttttccacc | gtcatcacccg | aaacgcgcga | gacgaaaggg | 2460 |
| cctcgtata | cgccatatttt | tataggttta | tgtcatgata | ataatggttt | cttagacgtc | 2520 |
| aggtggact | tttcggggaa | atgtgcgcgg | aacccttatt | tgtttat | tctaaataca | 2580 |
| ttcaaatatgc | tatccgctca | tgagacaata | accctgataa | atgcttcaat | aatattgaaa | 2640 |
| aaggaagagt | atgagtttcc | aacatttccg | tgtccgcctt | atccctttt | ttgcggcatt | 2700 |
| ttgccttcct | gtttttgtctc | acccagaaac | gctggtgaaa | gtaaaagatg | ctgaagatca | 2760 |

-continued

| | |
|--|------|
| gttgggtgca cgagtgggtt acatcgaaact ggatctcaac agcggtaaga tccttgagag | 2820 |
| tttcgcgccca gaagaacgtt ttccaatgtat gggactttt aaagttctgc tatgtggcgc | 2880 |
| ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac actattctca | 2940 |
| gaatgacttg gttgagactt caccagtac agaaaagcat cttacggatg gcatgacagt | 3000 |
| aagagaatta tgcgtgtct ccataaccat gagtgataac actgcggcca acttacttct | 3060 |
| gacaacgatc ggaggaccga aggagctaac cgctttttt cacaacatgg gggatcatgt | 3120 |
| aactcgccctt gatcggttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga | 3180 |
| caccacgatc cctgttagcaa tggcaacaac gttgcgcaaa ctattaactg gcgactact | 3240 |
| tactctagct tccccggcaac aattaataga ctggatggag gcccggataaag ttgcaggacc | 3300 |
| acttctgcgc tcggcccttc cggctggctg gtttattgtc gataaatctg gagccgggtga | 3360 |
| gcgtgggtct cgccgtatca ttgcagcact gggccagat ggttaagccct cccgtatcgt | 3420 |
| agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga | 3480 |
| gatagggtgcc tcactgatta agcattggta actgtcagac caagttact catatatact | 3540 |
| ttagattgtat taaaacttc attttaatt taaaaggatc taggtgaaga tccttttga | 3600 |
| taatctcatg accaaaaatcc cttAACGTGA gtttcgttc cactgagcgt cagaccgggt | 3660 |
| agaaaagatc aaaggatctt ctttagatcc ttttttctg cgcgtaatct gctgcttgca | 3720 |
| aacaaaaaaaaa ccaccgctac cagcgggtgtt ttgtttccg gatcaagagc taccaactct | 3780 |
| tttccgaag gtaactggct tcagcagacg gcagataccca aatactgtcc ttctagtgta | 3840 |
| gccgtgttta gccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgt | 3900 |
| aatcctgtta ccagtggctg ctgccagtg cgataagtgc tgtcttaccg ggttggactc | 3960 |
| aagacgatag ttacccgata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca | 4020 |
| gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga | 4080 |
| aagcgcacag cttcccgaaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg | 4140 |
| aacaggagag cgcacgaggag agcttccagg gggaaacgcc tggtatctt atagtcctgt | 4200 |
| cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtag gggggcggag | 4260 |
| cctatggaaa aacgcacgca acgcggcctt ttacggttc ctggccttt gctggccttt | 4320 |
| tgctcacatg gctcgacaga tcttcaatat tggccattag ccatattatt cattggttat | 4380 |
| atagcataaa tcaatattgg ctattggcca ttgcatacgt tggatctata tcataatatg | 4440 |
| tacatttata ttggctcatg tccaatatga ccgcctatgtt ggcattgatt attgactagt | 4500 |
| tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgg gttccgcgtt | 4560 |
| acataactta cggtaaatgg cccgcctggc tgaccgcccc acgacccccc cccattgacg | 4620 |
| tcaataatga cgtatgttcc catagtaacg ccaataggaa cttccattg acgtcaatgg | 4680 |
| gtggagtatt tacggtaaac tgccacttg gcaatcatac aagtgtatca tatggcaagt | 4740 |
| ccggcccccta ttgacgtcaa tgacggtaaa tggccgcctt ggcattatgc ccagtacatg | 4800 |
| accttacggg actttcctac ttggcagttac atctacgtat tagtcatcgc tattaccatg | 4860 |
| gtgatgcgggt tttggcagta caccaatggg cgtggatagc ggtttgactc acggggattt | 4920 |
| ccaagtctcc accccattga cgtcaatggg agttttttt ggcaccaaaa tcaacggac | 4980 |
| tttccaaat gtcgtaacaa ctgcgtatcgc ccggcccggt gacgcaatg ggcggtaggc | 5040 |
| gtgtacggtg ggaggcttat ataaggcagag ctgcgttagt gaaccgtcag atcactagaa | 5100 |
| gttattgtc ggtagttat cacagttaaa ttgctaacgc agtcaatgtctgacacaa | 5160 |

-continued

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| cagtcctcgaa | cttaagctgc | agtgactctc | ttaaggtagc | cttgcaagaag | ttggtcgtga | 5220 |
| ggcaactgggc | aggttaagtat | caagggttaca | agacagggtt | aaggagacca | atagaaactg | 5280 |
| ggcttgcga | gacagagaag | actcttgcgt | ttctgtatagg | cacctattgg | tcttactgac | 5340 |
| atccacttg | cctttctctc | cacaggtgtc | cactcccagt | tcaattacag | ctcttaaggc | 5400 |
| tagagtactt | aatacactc | actataggct | agcacatect | cttggtccta | tcacggttat | 5460 |
| gaggtcgacc | agttgttgct | ttgatgttcg | gttctctcg | tgattggac | aatatttggg | 5520 |
| geacttcgcc | ggtcccact | tccagaattt | ccgtgtggtc | tgtgaattta | tcacggctac | 5580 |
| actgtcatca | tattccagtt | ttgcaatctg | ctctctttgt | acctgcagat | aggtaccaa | 5640 |
| caaagttggg | taaggatagt | tcaatcaatg | atcattttct | agtgcactta | ggattcaaga | 5700 |
| tccttattatc | agggacaaga | gcaggattaa | ggatatccga | gtcgcgacgc | gtacatgtag | 5760 |
| cgctcgacc | ggtccgcggg | gctcgccctc | gaggtgcgag | aggccgagga | ccagaacaac | 5820 |
| atccgectac | ctccatcat | tgttataaaa | aacttaggaa | ccaggtccac | acagccgcca | 5880 |
| geccatcaac | catccactcc | cacgattgga | gccgcacgtg | tctagagggc | ccgtttaaac | 5940 |
| cctgcaggtt | taattaagt | aattcttgg | tgaactccgg | aaccctaatac | ctgcctctagg | 6000 |
| tggtaggca | ttatggca | tagattaaag | aaaacttga | aaatacgaag | tttctattcc | 6060 |
| cagcttgc | tggttttttt | cccccccaac | tccggaggtc | gaccagtact | ccggggcgaca | 6120 |
| ctttgtttt | ttttttccc | ccgatgctgg | aggtcgacca | gatgtccgaa | agtgtcccc | 6180 |
| cccccccccc | cccccccccc | gcgeggagcg | gccccccac | ccgggacccc | ttttttttt | 6240 |
| ttttttttt | ttttaattc | ctggAACCTT | taggtcgacc | agttgtccgt | ctttactcc | 6300 |
| ttcatatagg | tcgaccagta | ctccgggtgg | tactttgtct | ttttctgaaa | atcccagagg | 6360 |
| tcgaccagat | atccgc | | | | | 6376 |

<210> SEQ ID NO 6
<211> LENGTH: 3605
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 6

| | | | | | | |
|------------|-------------|-------------|------------|-------------|--------------|-----|
| agcttggcgt | aatcatggtc | atagctgttt | cctgtgtgaa | attgttatcc | gctcacaatt | 60 |
| ccacacaaca | tacgagccgg | aagcataaaag | tgtaaagcc | gggggtgccta | atgagtgcgc | 120 |
| taactcacat | taattgcgtt | gctgcactg | cccgcttcc | agtcgggaaa | cctgtcggtc | 180 |
| cagctgcatt | aatgaatcgg | ccaaacgcgcg | gggagaggcg | gtttgcgtat | tgggcgcct | 240 |
| tccgcttcc | cgctcactga | ctcgctgcgc | tccggcgttc | ggctgcggcg | agcggtatca | 300 |
| gtcactcaa | aggcggtaat | acggttatcc | acagaatcag | gggataacgc | aggaaagaac | 360 |
| atgtgagcaa | aaggccagca | aaaggccagg | aaccgtaaaa | aggccgcgtt | gttggcggtt | 420 |
| ttccataggc | tccggccccc | tgacgagcat | cacaaaaatc | gacgctcaag | tcaagagggtgg | 480 |
| cgaaacccga | caggactata | aagataaccag | gcgttcccc | ctggaaagctc | cctcggtcgc | 540 |
| tctcctgttc | cgaccctgcc | gcttaccgga | tacctgtccg | cctttctccc | ttcgggaagc | 600 |
| gtggcggtt | ctcatagctc | acgctgttagg | tatctcagg | cggtgttaggt | cgttcgtcc | 660 |
| aagctgggct | gtgtgcacga | accccccgtt | cagccgcacc | gctgcgcctt | atccggtaac | 720 |
| tatcgcttt | agtccaaaccc | ggtaagacac | gacttatcgc | cactggcagc | agccactgg | 780 |

-continued

-continued

| | | | | | | |
|------------|------------|-------------|-------------|------------|-------------|------|
| tccaatcggt | ggagtggatg | gttgcgtggc | tggcggtgt | gtggacctgg | ttccataagtt | 3180 |
| ttttataaca | atgtggagg | gttaggcggat | gttggttctgg | tcctcgccct | ctcgacactc | 3240 |
| gaggggcgcc | cccgeggacc | ggtgcgagcg | ctacatgtac | gcgtcgccac | tcggatatcc | 3300 |
| ttaatcgtc | tcttgtccct | gataatagga | tcttgaatcc | taagtgcact | agaaaatgat | 3360 |
| cattgattga | actatccta | cccaactttg | tttggtaacct | atctgcagg | acaaagagag | 3420 |
| cagattgcaa | aactgaaata | tgtacgtacgt | gtagcgggtg | taaattcaca | gaccacacgg | 3480 |
| aaattctgga | agtccggacc | ggcgaagtgc | cccaaataatt | gtcccaatca | acgagagaac | 3540 |
| cgaacatcaa | agcaacaact | ggtgcaccc | ataaccgtga | taggaccaag | aggatgtgct | 3600 |
| agccaa | | | | | | 3605 |

<210> SEQ_ID NO 7
<211> LENGTH: 3260
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 7

| | | | | | | |
|--------------|-------------|-------------|------------|-------------|-------------|------|
| ctctctcgcg | cggttcggtg | atgacgggtga | aaacctctga | cacatgcagg | tcccgagac | 60 |
| ggtcacagct | tgtctgtaa | cggtgcggg | gagcagacaa | gccccgtcagg | gcgcgtcagg | 120 |
| gggtgttggc | gggtgtcggt | gctggcttaa | ctatgcggca | tcagagcaga | ttgtactgag | 180 |
| agtgcacccat | atgcgggtgt | aaataccgca | cagatgcgt | aggagaaaat | accgcattcag | 240 |
| gcccattcg | ccattcaggc | tgcgcaactg | ttgggaagg | cgatcggtgc | gggccttcc | 300 |
| gctattacgc | cagctggcga | aagggggat | tgctgcaagg | cgattaagtt | gggttaacgcc | 360 |
| agggtttcc | cagtcacgac | gttgtaaaac | gacggccag | gaattcaagc | ttgctagcac | 420 |
| caactttgtt | ttgtctgtat | agtcgcgt | gacgaaaccc | ggagtcccg | gtcaccaaac | 480 |
| aaagttgggt | aaggatagtt | caatcaatga | tcattttcta | gtgcacttag | gattcaagat | 540 |
| cctattatca | gggacaagag | caggattaag | gatatccgag | tgcgcacgcg | tacatgtac | 600 |
| gctcgacccg | gtcccgcccc | cgcgcggcg | cgcctcgag | gtgcgagagg | ccgaggacca | 660 |
| gaacaacatc | cgccctaccct | ccatcattgt | tataaaaaac | ttaggaacca | ggtccacaca | 720 |
| gccgccagcc | catcaaccat | ccactcccac | gattggagcc | gcacgtgtct | agagggcccg | 780 |
| tttaaaccct | gcaggtttaa | ttaagtgtat | tcttgggtga | actccggAAC | cctaattctg | 840 |
| cccttaggtgg | ttaggcatta | tttgcata | attaaagaaa | actttgaaaa | tacgaagtt | 900 |
| ctatccccag | ctttgtctgg | tgccggccat | ggtcccagcc | tcctcgctgg | cgccgggtgg | 960 |
| gcaacattcc | gaggggaccg | tcccctcggt | aatggcgaat | gggaccgcgg | ccgcgcgc | 1020 |
| aagcttggcg | taatcatgg | catagctgtt | tcctgtgtga | aattgttatac | cgctcacaat | 1080 |
| tccacacacaac | atacgagccg | gaagcataaa | gtgtaaagcc | tggggtgcc | aatgagttag | 1140 |
| ctaactcaca | ttaattgcgt | tgcgcact | gcccgtttc | cagtcgggaa | acctgtcg | 1200 |
| ccagctgcgt | taatgaatcg | gccaacgcgc | ggggagaggc | ggtttgcgta | ttgggcgtc | 1260 |
| ttccgcgttcc | tgcgtcact | actcgctcg | ctcggtcg | cggctcgcc | gagcggtatc | 1320 |
| agctcactca | aaggcggtaa | tacggttatc | cacagaatca | ggggataacg | cagggaaagaa | 1380 |
| catgtgagca | aaaggccagc | aaaaggccag | gaaccgtaaa | aaggccgcgt | tgctggcg | 1440 |
| ttccatagg | ctccggcccc | ctgacgagca | tcacaaaaat | cgacgctcaa | gtcagaggt | 1500 |

-continued

| | |
|--|------|
| gcgaaacccg acaggactat aaagatacca ggcgttccc cctggaaagt ccctcggtcg | 1560 |
| ctctcctgtt ccgaccctgc cgcttacccg atacctgtcc gcctttctcc ctteggaaag | 1620 |
| cgtggcgctt ttcatacgct cacgctgttagt gtatctcagt tcgggttagg tcgttcgtc | 1680 |
| caagctgggc tgtgtgcacg aaccccccgt tcagccccacg cgctgegcct tatccggtaa | 1740 |
| ctatcgctt gagtccaacc cggtaagaca cgacttatcg ccactggcag cagccactgg | 1800 |
| taacaggatt agcagagcga ggtatgttagg cggtgctaca gagttttga agtggtgcc | 1860 |
| taactacggc tacactagaa gaacagtatt tggtatctgc gctctgtga agccagttac | 1920 |
| cttcggaaaa agagttggta gctcttgatc cggcaaacaaa accaccgcgtg gtacgggtgg | 1980 |
| ttttttgtt tgcaaggcgc agattacgcg cagaaaaaaaaa ggatctcaag aagatccccc | 2040 |
| gatctttctt acggggctcg acgctcgtg gaacgaaaac tcacgttaag ggatttttgtt | 2100 |
| catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttaa | 2160 |
| atcaatctaa agtataatgt agttaaacttg gtctgacagt taccaatgtct taatcagtga | 2220 |
| ggcacctata tcagcgatct gtctatccgtt ttcatccata gttgcctgac tccccgtcgt | 2280 |
| gtagataact acgatacggg agggcttacc atctggcccc agtgcgtcaca tgataccgcg | 2340 |
| agacccacgc tcaccggctc cagatttac agcaataaac cagccagccg gaaggccgaa | 2400 |
| gcgcagaagt ggtccgtcaa ctttatccgc ctccatccag tctattaaatt gttgccggaa | 2460 |
| agctagatgttgc cagttatag tttgcgcac gttgttgcac ttgctacagg | 2520 |
| catcgtggt tcacgtcgt cgtttgtat ggcttcattt agctccgggtt cccaaacgtatc | 2580 |
| aaggcggatc acatgatccc ccatgttgcg caaaaaagcg gttagtcct tcgggtcc | 2640 |
| gatcgttgcg agaagtaagt tggccgcagt gttatcactc atgggttatgg cagcaactgca | 2700 |
| taattctctt actgtcatgc catccgtaaatg atgctttctt gtgactgggt agtactcaac | 2760 |
| caagtcatcc tgagaatagt gtatgcggcg accgagttgc tcttgcggcg cgtcaatacg | 2820 |
| ggataataacc gcgccacata gcagaacttt aaaagtgcctc atcattggaa aacgttcc | 2880 |
| ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aacccactcg | 2940 |
| tgcacccaaac tgatcttcag catcttttac tttcaccaggc gtttctgggt gagcaaaaaac | 3000 |
| aggaaggcaa aatgcgcgaa aaaaggaaat aaggggcaca cggaaatgtt gaataactcat | 3060 |
| actcttcctt tttcaatattt attgaagcat ttatcagggt tattgtctca tgagcggata | 3120 |
| catatggaa tttttttttttttaaaaataaca aatagggtt ccgcgcacat ttccccggaaa | 3180 |
| agtgcacccat gacgtctaaat aaaccattat tatcatgaca ttaacctataaaaataggcg | 3240 |
| tatcacgagg cccttcgtc | 3260 |

<210> SEQ ID NO 8
 <211> LENGTH: 12844
 <212> TYPE: DNA
 <213> ORGANISM: artificial sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: sequence assembled from sequences obtained from
 different organisms or plasmids

<400> SEQUENCE: 8

| | |
|--|-----|
| tcgacgatat ctccagagga tcataatcg ccataccaca tttgttagagg ttttactgc | 60 |
| tttaaaaaac ctccccacacc tccccctgaa cctgaaacat aaaatgaatg caattgtgt | 120 |
| tgttaacttg tttattgcag cttataatgg ttacaataaa agcaatagca tcacaatattt | 180 |
| cacaaataaa gcattttttt cactgccccg agttcctcg ctcactgact cgctgcgc | 240 |

-continued

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|--------------|------|
| ggtcgttgcg | ctgcggcgag | cggtatcagc | tcactcaaag | gggtaataac | ggttatccac | 300 |
| agaatcagg | gataacgcag | gaaagaacat | gtgagcaaaa | ggccagcaaa | aggccaggaa | 360 |
| ccgtaaaaag | ggcgcttgc | tggegtttt | ccataggtctc | cgccccctg | acgagcatca | 420 |
| caaaaatcga | cgctcaagtc | agagggtggcg | aaaccgcaca | ggactataaa | gataccaggc | 480 |
| gttccccct | ggaagctccc | tcgtgcgtc | tcctgttccg | accctgcgc | ttaccggata | 540 |
| cctgtccgccc | tttctecctt | cggyaagcgt | ggcgctttt | caatgtcac | gctgttaggt | 600 |
| tctcagttcg | tgtaggtcg | ttcgctccaa | gctgggctgt | gtgcacgaac | cccccggtca | 660 |
| gccccgaccgc | tgcgccattat | ccggtaacta | tcgtcttgc | tccaacccgg | taagacacga | 720 |
| cttatacgcca | ctggcagcag | ccactggtaa | caggattagc | agagcgaggt | atgtaggcgg | 780 |
| tgctacagag | ttcttgaagt | ggtggctaa | ctacggctac | actagaagga | cagtatttgg | 840 |
| tatctgcgt | ctgctgaagc | cagttacctt | cggaaaaaga | gttggtagct | cttgatccgg | 900 |
| caaacaacc | accgctggta | gccccggg | ttttgttgc | aagcagcaga | ttacgcgcag | 960 |
| aaaaaaaaagg | tctcaagaag | atcctttgt | ctttctacg | gggtctgacg | ctcagtgaa | 1020 |
| cgaaaaactca | cggttaaggga | ttttggtcat | gagattatca | aaaaggatct | tcacccatag | 1080 |
| ccttttaaat | taaaaatgaa | gttttaatc | aatctaaagt | atatatgagt | aaacttggtc | 1140 |
| tgacagttac | caatgtttaa | tcagtggagc | acctatctca | cgatctgtc | tatttgcgtt | 1200 |
| atccatagtt | gcctgactcc | ccgtcggtgt | gataactacg | atacgggagg | gcttaccatc | 1260 |
| tggccccagt | gtgcataatga | taccgcgaga | cccacgctca | ccggctccag | atttatcagc | 1320 |
| aataaaccag | ccagccggaa | gggcggagcg | cagaagtgg | cctgcaactt | tatccgcctc | 1380 |
| catccagtc | attaatttgtt | gccccggaa | tagagtaagt | agttccgcag | ttaatagttt | 1440 |
| gcfgaacgtt | gttgcatttgc | ctacaggcat | cgtgggtgtca | cgctcgctgt | ttggatggc | 1500 |
| tttatttcgc | tcgggttccc | aacgtatcaag | gctgatgtaca | tgtatccccca | tgttgtgca | 1560 |
| aaaagcggtt | agtccttcg | gtcctccat | cgttgcaga | agtaagttgg | ccgcgtgttt | 1620 |
| atcaactcatg | tttatggcag | cactgcataa | tttcttctact | gtcatgcct | ccgtaagatg | 1680 |
| cttttctgtt | actggtgagt | actcaaccaa | gtcattctga | gaatagtgt | tgccggcacc | 1740 |
| gagttgtctt | tgccggcggt | caatacgggaa | taataccgc | ccacatagca | gaactttaaa | 1800 |
| agtgcgtatc | atggaaaac | gttcttcggg | gcaaaaactc | tcaaggatct | taccgtgtt | 1860 |
| gagatccagt | tgcgtatcac | ccactcggtc | acccaactga | tettcagcat | cttttacttt | 1920 |
| caccagcggt | tctgggtgag | aaaaacagg | aaggcaaaaat | gccgcaaaaa | agggaaataag | 1980 |
| ggcgacacgg | aaatgttga | tactcataact | cttcctttt | caatattatt | gaagcattta | 2040 |
| tcagggttat | tgtctcatga | gcccatacat | atttgaatgt | atttagaaaa | ataaacaataat | 2100 |
| aggggttccg | cgcacatttc | cccgaaaagt | gccacctgac | gtccggcagt | aaaaaaatgc | 2160 |
| tttatttgcgt | aaatttgtga | tgctattgt | ttatttgc | ccattataag | ctgcaataaa | 2220 |
| caagttaaaca | acaacaatgt | cattcatttt | atgtttcagg | tccaggggaa | ggtgtggag | 2280 |
| gttttttaaa | gcaagtaaaa | cctctacaaa | tgtggatgg | ctgattatga | tcctctagac | 2340 |
| tgcagactag | tttcgaaacta | gtctagaaga | tttctgtcat | tgtacactat | aggggtgtcc | 2400 |
| gtgtctgagc | cttgcgttcc | cgagattct | gccatggctt | gcagcctaag | caggcgctca | 2460 |
| gctgacccctc | gactgtccgt | cggtatctgg | ctggactccg | atgcagtg | aatgtctagg | 2520 |
| ggtgtgcggg | ttggaagatg | ggcagctctc | gcatcaactt | ctctgttggg | cccggtttct | 2580 |
| ctgtaqctct | ccctqgctc | tcctcqactc | tqtttqaccc | tcctatcttc | tttqcccccc | 2640 |

-continued

| | |
|---|------|
| aatctcggt a gtcattctc actttgatca ccgtgtagaa atgataactg ggcttgc tctg | 2700 |
| ggtccaa c cg ctctactgat cttgtcctca gtagtatgca ttgcaatctc tgaaaaca a g c | 2760 |
| cttgcatctt c cggcagt gat acccgat tca gat gccaat g tggaactgac cttccagct | 2820 |
| gacctcc tta ccatctctt cccta atctt aaatatgctg gat caa agta agat cggcca | 2880 |
| aagg tcaaa ac ctccc atgga gtttcaagt tccacttca ctccc atgga atagctccag | 2940 |
| agcagagggt atgatcc tgc actgaa ctg ttctgaa attg agttctccag gattaccatg | 3000 |
| taggg tgcag tttccccat ttgctggtaa aggttcatca aggactcaag tggataac | 3060 |
| tcaccagcaa attcatgcag tcca a gaga gca ggatacatag tttctatccc aaacttaata | 3120 |
| gtcaggataa aactggctaa tcctgcctt acgatata tgc tatcaatgtc acatatcatt | 3180 |
| tcagcaatcc tggg tttgtt tccgggtt ctctt gatat ccaggattag agcgaccatg | 3240 |
| aatcggcgt a aggagaggc ctcggcaatc ctgttca ccacatccaa ccattttctc | 3300 |
| tccaa tctaa attcaccaac tacccttctt ttttttttgc ac tttatcca ctttcttagc | 3360 |
| tccgaatcag otgcccgtgc tggggccgt a accgcctt cgagcaagac ccaaatttga | 3420 |
| gtttagatgg tacc cagaat catgtt gat ccctcagggt ctgcacttca aatatctgag | 3480 |
| at ttc ttgt ttc tca acca tccgaacctg gattgatc ac ttaatttgg atcatcatgt | 3540 |
| aaaaa gtattt ggtccgcctc atcctccatg ttgttacctc ttgtatgc gaa ggtaaggcca | 3600 |
| gattgtgact ggtcactctg gacaacctt aacagcctt tgc taa cgtc aggg tcatcg | 3660 |
| gtgatectct gaatcaattt acctggagac tccaca aata aggataat accttatttgc | 3720 |
| gcccctgtta gtttggccc gtcacatcc gggtttccaa ttaacctgac caacgg tcc | 3780 |
| agaagtctgg atcgagtgg aatttggaa tctccaggaa ttgg tttactat aataatgtgt | 3840 |
| ttgatttctc t gatggcttcc accggatct gat gat tgg gtttgc ctttttctt | 3900 |
| ttgaacaatg ctaa gctt taaaatgtt gcat gat tccaggcga tctgac gtt | 3960 |
| cactaaacga gctctgctt tata ggc tcc acccg taca cggccac tctc acatactcg | 4020 |
| gtatgttata atagtaatca attacggggtt cattatgttca tagccat atggat tcc | 4080 |
| gcgttacata acttacggta aatggccgc ctggctgacc gccaacgac cccgccc at | 4140 |
| tgacgtcaat aatgacgtat gttccatag taac gccaat agggactt cattgac gtc | 4200 |
| aatgggtt gat ttttacgg taaaactgccc acttggcagt acatcaatg ttttcatatgc | 4260 |
| caagtacgcc ccctatttgc gtc aatgac gtc gtaa atggcc gtc cccctggcat tatttgc | 4320 |
| acatgacctt atggactttt cctacttggc agtacatctt cgtatttgc atcgat ttttgc | 4380 |
| ccatggtgat gcggttttgg cag taca tca atggcgtt gat ggg ttttgc gactc ac | 4440 |
| gatttccaa g tctccacccc attgacgtca atgggagttt gtttggc ac caaaatcaac | 4500 |
| gggactttcc aaaatgtcgt aacaactccg ccccat gac gcaatggc ggtaggcgt | 4560 |
| tacgggtggaa ggtctata a gca gat gtc gttttagt gaa ccgtc a gat gtc | 4620 |
| cctctagtc a gtc a gtc | 4680 |
| gactggaatg ctc a cccat gtc | 4740 |
| ctatggc a gtc | 4800 |
| aaggagaaggc aggca gat ttc ggtctc acaat gtc gtc gtc gtc gtc gtc gtc | 4860 |
| aaggcggtgc acctcgc acaat gtc | 4920 |
| cttggaaat ccccaaga aatctccagg catcaagcac tgggttacag ttttattatg | 4980 |

-continued

| | |
|--|------|
| tttatgatca cagcggtgaa gcggttaagg gaatccaaga tgctgactct atcatggttc | 5040 |
| aatcaggcct ttagtggatg agcaccctct caggaggaga caatgaatct gaaaacagcg | 5100 |
| atgtggatat tggcgaaacct gataccgagg gatatgctat cactgaccgg ggatctgctc | 5160 |
| ccatctctat ggggttcagg gcttctgtat ttgaaactgc agaaggaggg gagatccacg | 5220 |
| agctccttag agtccaaatcc agaggcaaca actttccgaa gcttgggaaa actctcaatg | 5280 |
| tccctccgccc tccggacccc ggttagggca gcacttccgg gacaccatt aaaaaggcga | 5340 |
| cagacgcgag attagcctca tttgaaacgg agatcgcgtc tttattgaca ggtggtcaa | 5400 |
| cccaatgtgc tcgaaagtca ccctcggAAC catcaggAAC aggtgcacct gcggggaaatg | 5460 |
| tccccgagtg tgtgagcaat gccgcactga tacaggagtg gacaccggaa tctggtagcca | 5520 |
| caatctcccc gagatcccg aataatgaag aaaaaaaaaaaaa ctattatgtat gatgagctgt | 5580 |
| tctctgatgt ccaagatatt aaaacagcct tggccaaat acacgaggat aatcagaaga | 5640 |
| taatctccaa gctagaatca ctgctgttat tgaaggaga agttgagtc attaagaagc | 5700 |
| agatcaacag gcaaaatatac agcatatcca ccctggAAAG acacctctca agcatcatga | 5760 |
| tcgcccattcc tggacttggg aaggatccca acgacccccc tgcagatgtc gaaatcaatc | 5820 |
| ccgacttgaa accccatcata ggcagagatt caggccgagc actggccgaa gttctcaaga | 5880 |
| aacccgttgc cagccgacaa ctccaaaggaa tgacaaatgg acggaccagt tccagaggac | 5940 |
| agctgctgaa ggaatttcag ctaaagccga tcggggaaaaa gatgagtc gccgtcggt | 6000 |
| ttgttctgtc caccggccct gcacacgcgca gtgtaatccg ctccattata aaatccagcc | 6060 |
| ggctagagga ggatcggaaag cgttacctga tgactctcct tgatgatatc aaaggagcca | 6120 |
| atgatcttgc caagttccac cagatgctga tgaagataat aatgaagtag cggccgttct | 6180 |
| gacatccggc gggttctga catccggcg gttctgaca tccgggggt ttctgacatc | 6240 |
| cggccggttt ctgacatccg cgggtgtact cacaacggat ccaacagaca tatggactcg | 6300 |
| ctatctgtca accagatctt ataccctgaa gttcacctag atagcccgat agttaccaat | 6360 |
| aagatagtag ccacccctgga gtatgctga gtcctcactg cttacagcct ggaggaccct | 6420 |
| acactgtgtc agaacatcaa gcaccccta aaaaacggat ttccaaacca aatgattata | 6480 |
| aacaatgtgg aagttggaa tgcacatcaag tccaaagtttta ggagttatcc ggccactct | 6540 |
| catattccat atccaaattt taatcaggat ttatTTAACA tagaagacaa agatcaacg | 6600 |
| aggaagatcc gtgaactcct caaaaagggg aattcgctgt actccaaagt cagtgataag | 6660 |
| gttttccaaat gcttaaggga cactaactca cggcttggcc taggctccga attggggag | 6720 |
| gacatcaagg agaaagttt taacttggga gtttacatgc acagctccca gtggttttag | 6780 |
| ccctttctgt ttgggtttac agtcaagact gagatggggat cagtgattaa atcacaacc | 6840 |
| catacttgcc ataggaggag acacacaccc gtattctca ctggtagttc agttgagtt | 6900 |
| ctaatctctc gtgacccctgt tgctataatc agttaaagagt ctcaacatgt atattaccc | 6960 |
| acatttgaac tgggtttgtat gtattgtat gtcatagagg ggaggtaat gacagagacc | 7020 |
| gtatgacta ttgatgctag gtatcagag cttcttaggaa gagtcagata catgtggaaa | 7080 |
| ctgatagatg gtttctccc tgcactcggg aatccaaactt atcaaattgt agccatgt | 7140 |
| gaggcctttt cacttgctta cctgcagctg agggatataa cagtagaaact cagaggtgt | 7200 |
| tcccttaacc actgttttac tggaaatacat gatgttttg accaaaaacgg gtttctgtat | 7260 |
| gaaggtagttt atcatgagtt aattgaagct ctagattaca ttttctaaac tgatgacata | 7320 |
| catctgacag gggagatttt ctcattttc agaagttcg gccaccccg acttgaagca | 7380 |

-continued

-continued

| | |
|--|-------|
| atgtgggcga ggctagctcg aggacggcct atttacggcc ttgaggtccc tgatgtacta | 9780 |
| gaatctatgc gagggcacct tattcggcgt catgagacat gtgtcatctg cgagtgtgga | 9840 |
| tcaagtcaact acggatggtt ttttgcctcc tcgggttgcc aactggatga tattgacaag | 9900 |
| gaaacatcat ccttgagagt cccatatatt ggttctaccac ctgatgagag aacagacatg | 9960 |
| aagcttcgcct tcgttaagagc cccaaatgtcgat tccttgcgtat ctgtgttag aatagcaaca | 10020 |
| gtgtactcat gggcttacgg tgatgtat agcttgcgtat agcaaggctg gttgtggct | 10080 |
| aggcaaaggc ccaatgtgag cctggaggag ctaagggtga tcactccat ctcaacttcg | 10140 |
| actaatttag cgcatgggtt gagggtatcgat agcactcaag tgaaataactc aggtacatcc | 10200 |
| cttgcgcag tggcgaggta taccacaatc tccaacgaca atctctcatt tgtcatatca | 10260 |
| gataagaagg ttgatactaa ctttatatac caacaaggaa tgcttctagg gttgggttt | 10320 |
| ttagaaacat tgttcgact cgagaaagat accggatcat ctaacacggtt attacatctt | 10380 |
| cacgtcgaaa cagattgttg cgtgtatcccg atgatagatc atcccaggat acccagctcc | 10440 |
| cgcgaagctg agctgagggc agagctatgt accaaccat tgatatatga taatgcacct | 10500 |
| ttaattgaca gagatacaac aaggctatac acccagagcc ataggaggca ccttgcggaa | 10560 |
| tttgcgttacat ggtccacacc ccaactatatac cacattttag ctaagtccac agcactatct | 10620 |
| atgattgacc tggtaacaaa atttgagaag gaccatatga atgaaatttc agctctata | 10680 |
| ggggatgacg atatcaatag tttcataact gagtttgcgt tcatacgacc aagattttc | 10740 |
| actatctact tggccagtg tgccgcattc aattggccat ttgatgtaca ttatcataga | 10800 |
| ccatcaggaa aatatcgat gggtgagctg ttgtcatcgat tcctttctag aatgagcaaa | 10860 |
| ggagtgttta aggtgcttgtt caatgctcta agccacccaa agatctacaa gaaattctgg | 10920 |
| cattgtgttta ttatagagcc tatccatgtt ctttcacttg atgctcaaaa cttgcacaca | 10980 |
| actgtgtgca acatggttta cacatgtat atgacctacc tcgacctgtt gttgaatgaa | 11040 |
| gagtttagaaag agttcacatt tctttgtgtt gaaagcgcacg aggatgttagt accggacaga | 11100 |
| ttcgacaaca tccaggcaaa acacttatgt gttctggcag atttgtactg tcaaccagg | 11160 |
| acctgccccac caattcgagg tctaaagaccc gtagagaaat gtgcagttt aaccgaccat | 11220 |
| atcaaggcag aggctaggat atctccagca ggatcttcgtt ggaacataaa tccaattttt | 11280 |
| gtagaccatt actcatgctc tctgacttat ctccggcggat gatcgatcaa acagataaga | 11340 |
| ttgagagttt atccaggatt catttcgac gcccctcgat aggtaaatgt cagtcagcca | 11400 |
| aagatcggca gcaacaacat ctcaaatacg agcatcaagg ctttcgaccc cccacacgat | 11460 |
| gatgttgcac aattgtctaa agatatac acaagcaagc acaatctcc catttcaggg | 11520 |
| ggcaatctcg ccaattatga aatccatgtt ttccgcggaa tcgggttgcctt ctcatctgt | 11580 |
| tgctacaaag ctgttggat atcaacatta attaggat gcttgcggcc agggggaggac | 11640 |
| ggcttgcgtt tgggtgggg atcgggttcc atgttgcata cttataagga gatactaaaa | 11700 |
| ctaaacaagt gcttctataa tagtggggtt tcggccaaat ctgatctgg tcaaaggaa | 11760 |
| ttagcaccct atccctccga agttggcctt gtcgaacaca gaatggggat aggtaatatt | 11820 |
| gtcaaaagtgc tcttaacgg gaggccggaa gtcacgtggg taggcagtgtt agattgcctc | 11880 |
| aatttcatacg ttagtaataat ccctacatctt agtgtggggat ttatccattc agatatacg | 11940 |
| accttgcata cacaagatac tatagagaag cttagggat tggcagccat cttatcgatg | 12000 |
| gctctgcgtcc tgggcggaaat aggatcaata ctgggtgatca agcttgcgtt tttcagccgg | 12060 |
| gatttgttcc agggattttt aagttatgtt gggtcccatt atagagaagt gaaaccttgcata | 12120 |

-continued

| | |
|---|-------|
| taccctagat acagcaactt catabctact gaatcttatt tggttatgac agatctcaag | 12180 |
| gctaaccggc taatgaatcc tggaaagatt aagcagcaga taattgaatc atctgtgagg | 12240 |
| acttcacctg gacttatagg tcacatccta tccattaagc aactaagctg catacaagca | 12300 |
| attgtgggag acgcagttag tagaggtat atcaatccta ctctgaaaaa acttacacct | 12360 |
| atagagcagg tgctgatcaa ttgcgggtt gcaattaacg gacctaagct gtgcaaagaa | 12420 |
| ttgatccacc atgatgttgc ctcaggcaaa gatggattgc ttaattctat actcatcctc | 12480 |
| tacagggagt tggcaagatt caaagacaac caaagaagtc aacaaggat gttcacgct | 12540 |
| taccccgat tggtaagtag caggcaacga gaacttatat ctaggatcac ccgcaattt | 12600 |
| tggggcaca ttcttcttta ctccggaaac agaaagttga taaataagtt tatccagaat | 12660 |
| ctcaagtccg gctatctgat actagactta caccagaata tcttcgttaa gaatctatcc | 12720 |
| aagtcaaaaa aacagattat tatgacgggg ggtttgaac gtgagtgggt ttttaaggta | 12780 |
| acagtcaagg agaccaaaa atggtataag ttatcgat acagtgcct gattaaggac | 12840 |
| taag | 12844 |

<210> SEQ ID NO 9
<211> LENGTH: 15809
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence assembled from sequences obtained from different organisms or plasmids

<400> SEQUENCE: 9

| | |
|--|------|
| gcttcccttt agtgagggtt aatgcttcga gcagacatga taagatacat ttagtggat | 60 |
| ggacaaacca caactagaat gcagtggaaa aaatgctta ttgtgaaat ttgtgtatgc | 120 |
| attgctttat ttgttaaccat tataagctgc aataaaacaag ttaacaacaa caattgcatt | 180 |
| cattttatgt ttcaggttca gggggagatg tgggaggttt tttaaagcaa gtaaaacctc | 240 |
| tacaaatgtg gtaaaatccg ataaggatcg atccggctg gcttaatagc gaagaggccc | 300 |
| gcaccgatcg ccctcccaa cagttgcgcgc gcctgaatgg cgaatggacg cgccctgttag | 360 |
| cggcgcatta agcgcggcgg gtgtgggtt tacgcgcgcgt gtgaccgcata cacttgcag | 420 |
| cgccctagcg cccgctccctt tcgcctttt cccttcctt ctgcgcacgt tcgcccgtt | 480 |
| tccccgtcaa gctctaaatc gggggctccc tttagggttc cgattttagag cttaacggca | 540 |
| cctcgaccgc aaaaaacttg atttgggtga tggttcacgt agtggccat cgccctgata | 600 |
| gacggttttt cgccctttga cgttggagtc cacgttctt aatagtggac tcttggcca | 660 |
| aacttggaaaa acactcaacc ctatctcggt ctattctttt gatttataag ggattttgcc | 720 |
| gatttcggcc tattggtaa aaaatgagct gattnaacaat atatttaacg cgaattttaa | 780 |
| caaaatatta acgtttacaa ttgccttgc tgcggatattt ttccttacg catctgtgc | 840 |
| gtatttcaca ccgcatacgc ggatctgcgc agcaccatgg cctgaaataa cctctgaaag | 900 |
| aggaacttgg ttaggtacct tctgaggcgg aaagaaccag ctgtggaaatg tggatgtcgtt | 960 |
| agggtgtgga aagtccccag gctcccccagc aggccagaatg atgcaaaagca tgcatctcaa | 1020 |
| tttagtcagca accaggtgtg gaaagtcccc aggtccccca gcaggcagaa gtatgcaag | 1080 |
| catgcacatc aatttagtcag caaccatgtt cccggccctt actccgcocca tccggccct | 1140 |
| aactccgccc agttccgccc attctccgcctt ccatggctga ctaattttt ttatattgc | 1200 |
| agaggccgag gccgcctcggtt cctctgagctt attccagaag tagtggaggag gctttttgg | 1260 |

aggccttaggc ttttgaaaaa agcttggattc ttctgacaca acagtctcg aacttaaggct 1320
 agagccacca tgattgaaca agatggattt caccgggtt ctccggccgc ttgggtggag 1380
 aggctattcg gctatgactg ggcacaacag acaatcggtt gctctgatgc cgccgtttc 1440
 cggctgttcg cgcaggggcg cccggttttt tttgtcaaga ccgacctgtc cggtgccctg 1500
 aatgaactgc aggacgaggc agcgccgcta tcgtggctgg ccacgacggg cggtccctgc 1560
 gcagctgtgc tcgacgttgtt cactgaagcg ggaaggact ggctgttattt gggcgaatgt 1620
 cccggggcagg atctctgtc atctcacattt gctctgtccg agaaagtata catcatggct 1680
 gatgcaatgc ggccggctgca tacgcttgcgat cccggttccat gcccattcga ccaccaagcg 1740
 aaacatcgca tcgagcgagc acgtactcgat atggaagecg gtcttgcga tcaggatgt 1800
 ctggacaaag agcatcaggc gctcgccca gccgaactgtt tggccaggctt caaggccgc 1860
 atgccccacg gcgaggatctt cgtcgatccat gatggcgatg cctgcttgcga gaatatcatg 1920
 gtggaaaatg gccgttttc tggatttgcgact gactgtggccg ggctgggtgtt ggcggaccgc 1980
 tatcaggaca tagcggttgc taccgtgtat attgctgaag agcttggccg cgaatggct 2040
 gaccgttcc tcgtgtttt cggatcgcc gctccggattt cgcagecgat cgccttctat 2100
 cgccttcttgc acgagttttt ctgagcgccgat ctctgggtt cggaaatgacc gaccaaggcg 2160
 cgcacccaccc gccatcacga tggccgcaat aaaatatctt tattttcattt acatctgtgt 2220
 gtgggtttt tggatgttgcgata aggttcccgat tatggtgacac tctcagttaca 2280
 atctgtctgc atgcgcata gttaaaggccg ccccgacacc cgcacccaccc cgctgacgcg 2340
 ccctgacggg ctgtgtgtt cccggatcc gcttacagac aagctgtgac cgtctccggg 2400
 agctgtatgtt gtcaggggtt ttaccgtca tcaccggaaac ggcggacacgg aaaggccctc 2460
 gtgatacgcc tattttata ggttaatgtc atgataataa tggtttctt gacgtcagg 2520
 ggcacttttcc gggaaatgtt ggcggaaacc cctattttgtt tattttctt aatacatca 2580
 aatatgtatc cgctcatgat acaataaccc tgataatgtc ttcaataata ttgaaaagg 2640
 aagagtatgtatc gtattcaaca ttccgtgtc gcccatttttcc ctttttttgc ggcattttgc 2700
 ctccctgtttt ttgctcaccc agaaacgctg gtgaaagttaa aagatgtgatc agatcagg 2760
 ggtgcacgg tgggttacat cgaactggat ctcaacagcg gtaagatctt tgagatgtt 2820
 cgcggccaaag aacgtttcc aatgtatgtc acttttaaag ttctgtatg tggcgccgt 2880
 ttatcccgta ttgacggccg gcaagagcaa ctgggtcgcc gcatacacta ttctcagaat 2940
 gacttgggtt gactactcacc agtcacagaa aagcatcttta cggatggcat gacagtaaga 3000
 gaattatgca gtgctccat aaccatgatc gataacactg cggccaaactt acttctgaca 3060
 acgatcggag gaccgaagga gctaaccgtt ttttgcaca acatggggga tcatgtact 3120
 cgccttgcgtt gttggaaacc ggagctgtatc gaagccatccaaacgacga ggcgtacacc 3180
 acgatgcctg tagcaatggc aacaacgttgc cgcaaaactat taactggcgacttact 3240
 ctatgttttttcc ggcaacaattt aatagactgg atggaggccg ataaagttgc aggaccactt 3300
 ctgcgtcgccg ccctccggc tgggtttttt attgctgtatc aatctggatc cggtgagcgt 3360
 ggggtctcgccg gtatcattgc agcactgggg ccagatggta agccctcccg ttcgttagtt 3420
 atctacacgca cggggaggatc ggcacactatg gatgaacgaa atagacagat cgctgagata 3480
 ggtgcctcactt gatgttactg tcagaccaag tttactcata tataacttttag 3540
 atgatttaa aacttcattt ttaattttaa aggttcttggg tgaagatctt ttttgcataat 3600

US 9,441,205 B2

49

50

-continued

| | |
|--|------|
| ctcatgacca aaatccctta acgtgagtt tcgttccact gagcgtcaga ccccgtagaa | 3660 |
| aagatcaaag gatcttcttg agatcctttt ttctcgccg taatctgcgtc cttgcaaaaca | 3720 |
| aaaaaaaccac cgctaccagg ggtggtttgt ttgcggatc aagagctacc aactctttt | 3780 |
| ccgaaggtaa ctgggttcag cagagcgcag ataccaaata ctgtccttct agtgttagccg | 3840 |
| tagtttaggcc accacttcaa gaactctgtc gcaccgccta catacctcgc tctgtcta | 3900 |
| ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga | 3960 |
| cgatagttac cggataaggc gcagcggcgtg ggctgaacgg ggggttcgtg cacacagccc | 4020 |
| agcttggcgc gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaagc | 4080 |
| gccacgcgttc ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca | 4140 |
| ggagagcgcg cggaggagct tccaggggaa aacgcctggg atctttatag tcctgtcggg | 4200 |
| tttcgcacc tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gggagccta | 4260 |
| tggaaaaacg ccagcaacgc ggcctttta cggttctgg cctttgcgtg gcctttgct | 4320 |
| cacatggcgc gacagatctt caaatattggc cattagccat attattcatt gtttatata | 4380 |
| cataaatcaa tattggctat tggccattgc atacggttgc tctatatcat aatatgtaca | 4440 |
| tttatattgg ctcatgtcca atatgaccgc catgttggca ttgattatg actagttatt | 4500 |
| aatagtaatc aattacgggg tcattagttc atagccata tatggagttc cgcgttacat | 4560 |
| aacttacggt aaatggcccg cctggctgac cgcccaacga ccccccggca ttgacgtcaa | 4620 |
| taatgacgta tgccccata gtaacgc当地 tagggacttt ccattgacgt caatgggtgg | 4680 |
| agtatttacg gtaaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtccgc | 4740 |
| cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct | 4800 |
| taacggactt tcctacttgg cagtagatct acgtatttgc catcgcttatt accatggta | 4860 |
| tgccgttttg gcagtagacacc aatgggcgtg gatagcggtt tgactcacgg ggattccaa | 4920 |
| gtctccaccc cattgacgta aatgggagtt tgttttggca caaaaatcaa cgggacttcc | 4980 |
| caaatgtcg taacaactgc gatcgccgc cccgttgcacg caaatggcg gtaggcgtgt | 5040 |
| acgggtggag gtctatataa gcagagctcg ttttagtgaac cgtagatca ctagaagctt | 5100 |
| tattgcggta gtttatcaca gttaaattgc taacgc当地 agtgc当地tgc acacaacagt | 5160 |
| ctcgaactta agctgc当地tgc actctctttaa ggtagccctt cagaagttgg tcgtgaggca | 5220 |
| ctggcaggta aagtatcaag gttacaagac aggttaagg agaccaatag aaactggct | 5280 |
| tgtcagacaca gagaagactc ttgc当地tgc gataggcacc tattggctt actgacatcc | 5340 |
| actttgc当地tgc tctctccaca ggtgtccact cccagttcaa ttacagctt taaggctaga | 5400 |
| gtacttaata cgactcaacta taggctagca tggccacact tttaaggagc tttagcttgc | 5460 |
| tcaaaaagaaa caaggacaaa ccacccatta catcaggatc cgggtggagcc atcagaggaa | 5520 |
| tcaaaacat tattatagta ccaatccctg gagattcctc aattaccact cgatccagac | 5580 |
| ttctggaccc gttggc当地tgc ttaattggaa acccggatgt gagcgggccc aaactaacag | 5640 |
| gggcactaat aggtatatta tccttatttg tggagtc当地tgc aggtcaatg attcagagga | 5700 |
| tcaccgatga ccctgacgatc agcataaggc tggtagaggt tggccagagt gaccagtcac | 5760 |
| aatctggctt taccttc当地tgc tcaagaggta ccaacatggaa ggtgaggcg gaccaatact | 5820 |
| tttcacatga tgatccaaatt agtagtgc当地tgc aatccaggtt cggatggc当地tgc gagaacaagg | 5880 |
| aaatctcaga tattgaagtg caagaccctg agggattcaa catgattctg ggtaccatcc | 5940 |
| tagctcaaatt tgggtcttgc ctcgcaaaagg cgggtacggc cccagacacgc gcaagctgatt | 6000 |

-continued

cggagctaag aagggtggata aagtacaccc aacaaagaag ggtagttggt gaattttagat 6060
 tggagagaaa atgggtggat gtggtgagga acaggattgc cgaggacctc tccttacgcc 6120
 gattcatggt cgctctaatac ctggatatca agagaacacc cggaaacaaa cccaggattg 6180
 ctgaaatgat atgtgacatt gatacatata tcgttagggc aggattagcc agtttatcc 6240
 tgactattaa gtttggata gaaactatgt atcctgtct tggactgcat gaatttgctg 6300
 gtgagttatc cacacttgag tccttgatga acctttacca gcaaattggg gaaactgcac 6360
 cctacatggt aatcctggag aactcaattc agaacaagt cagtgcagga tcataccctc 6420
 tgctctggag ctatgccatg ggagtagggag tggaaacttga aaactccatg ggaggttga 6480
 actttggccg atcttacttt gatccagcat atttttagatt agggcaagag atggtaagga 6540
 ggtcagctgg aaaggtcagt tccacattgg catctgaact cggatctact gccgaggatg 6600
 caaggcttgt ttcagagatt gcaatgcata ctactgagga caagatcagt agageggttg 6660
 gacccagaca agcccaagta tcatttctac acgggtgatca aagtgagaat gagctaccga 6720
 gattgggggg caaggaagat aggagggtca aacagagtcg aggagaagcc agggagagct 6780
 acagagaaac cgggeccagc agagcaagtg atgcgagagc tgcccatctt ccaaccggca 6840
 cacccttaga cattgacact gcattcgagt ccagccaaga tccgcaggac agtgcaggt 6900
 cagctgacgc octgettagg ctgcaagcca tggcaggaat ctggaaagaa caaggctcag 6960
 acacggacac ccctatactg tacaatgaca gaaatcttct agacggctcc ggagccacga 7020
 acttctctct gttaaagcaa gcaggagacg tggaaagaaaa cccccgtccc atggcagaag 7080
 agcaggcacg ccatgtcaaa aacggactgg aatgcacccg ggctctcaag gccgagccca 7140
 tggcgtcaact ggcacatcgag gaagctatgg cagcatggc agaaatatac gacaaccag 7200
 gacaggagcg agccacactgc agggaaagaga aggcaggcag ttccgggtctc agcaaaccat 7260
 gcctctcagc aattggatca actgaaggcg gtgcaccccg catccgcggc cagggacactg 7320
 gagagagcga tgacgacgct gaaactttgg gaatcccccc aagaaatctc caggeatcaa 7380
 gcactgggtt acagtgttat tatgtttatg atcacagcgg tgaagegggtt aagggaatcc 7440
 aagatgctga ctctatcatg gttcaatcag gccttgatgg tgatagcacc ctctcaggag 7500
 gagacaatga atctgaaaac agcgatgtgg atattggcga acctgtatcc gagggatatg 7560
 ctatcactga cggggatct gctcccatct ctatgggtt cagggcttct gatgtgaaa 7620
 ctgcagaagg agggggatcc cacgactcc tgagactcca atccagagcc aacaacttc 7680
 cgaagcttgg gaaaactctc aatgttcctc cgccctccggc ccccggtagg gccagactt 7740
 cccggacacc cattaaaaag ggcacagacg cgagattacg ctcatggc acggagatcg 7800
 cgtcttatt gacagggtggt gcaacccat gtgcgtgaaa gtcacccctcg gaaccatcag 7860
 ggcctggc acctgccccg aatgtccccg agtgtgtgag caatgcccgc ctgatacagg 7920
 agtggacacc cgaatctggt accacaatct ccccgagatc ccagaataat gaagaaggg 7980
 gagactatta tgatgtgag ctgttctctg atgtccaaga tattaaaaca gccttggcca 8040
 aaatacacgaa ggataatcag aagataatct ccaagctaga atcaactgctg ttattgaagg 8100
 gagaagttga gtcaattaag aagcagatca acaggcaaa tatcagcata tccaccctgg 8160
 aaggacacct ctcaagcatc atgatcgcca ttccctggact tggaaaggat cccaaacgacc 8220
 ccactgcaga tgtcggaaatc aatcccgact tggaaacccat cataggcaga gattcaggcc 8280
 gagcactggc cgaagttctc aagaaacccg ttgcagccg acaactccaa ggaatgacaa 8340

-continued

| | |
|--|-------|
| atggacggac cagttccaga ggacagctgc tgaaggaatt tcagctaaag ccgatcggga | 8400 |
| aaaagatgag cttagccgtc gggtttggtc ctgacaccgg ccctgcatca cgcgatgtaa | 8460 |
| tccgctccat tataaaatcc agccggctag aggaggatcg gaagcgttac ctgatgactc | 8520 |
| tccttgatga tatcaaagga gccaatgatc ttgccaagtt ccaccagatg ctgatgaaga | 8580 |
| taataaatgaa gtagacgcgt cgagcatgca tctagggcgg ccaattccgc cccttcccc | 8640 |
| ccccccccctc tccctccccc cccctctaagc ttactggccg aagccgcttg gaataaggcc | 8700 |
| ggtgtgcgtt tgtctatatg tgatttcca ccatattgcc gtctttggc aatgtgaggg | 8760 |
| cccgaaacc tggccctgtc ttcttgacga gcattcttag gggctttcc cctctcgcca | 8820 |
| aaggaatgca aggtctgttg aatgtcgtga aggaagcagt tcctctggaa gcttcttga | 8880 |
| gacaaacaac gtctgtagec accctttgca ggcagcggaa ccccccacct ggcgacaggt | 8940 |
| gcctctgcgg caaaaagcca cgtgtataag atacacctgc aaaggcggca caacccagt | 9000 |
| gccacgttgt gagttggata gttgtggaaa gagtcaaatg gctctctca agcgtattca | 9060 |
| acaaggggct gaaggatgcc cagaaggatc cccattgtat gggatctgtat ctggggcctc | 9120 |
| ggtgcacatg ctttacatgt gtttagtcga ggttaaaaaaa acgtcttaggc ccccgaaacc | 9180 |
| acggggacgt ggtttccct tgaaaaacac gatgataagc ttgccacaac cggggatcct | 9240 |
| ctagagtcga catggactcg ctagtgcata accagatctt ataccctgaa gttcacctag | 9300 |
| atagccccat agttaccaat aagatagtag ccatcctgga gtatgtcga gtccctcagc | 9360 |
| cttacagcct ggaggaccct acactgtgtc agaacatcaa gcaccgccta aaaaacggat | 9420 |
| tttccaacca aatgattata aacaatgtgg aagttggaaa tgtcatcaag tccaaagctta | 9480 |
| ggagttatcc ggcacactct catattccat atccaaatgg taatcaggat ttatthaaca | 9540 |
| tagaagacaa agagtcaacg aggaagatcc gtgaactcct caaaaaggaa aattcgtgt | 9600 |
| actccaaagt cagtgataag gtttccaaat gcttaaggaa cactaactca cggcttggcc | 9660 |
| taggctccga attgagggag gacatcaagg agaaagttag taacttggaa gtttacatgc | 9720 |
| acagctccca gtggttttag cccttctgt tttggtttac agtcaagact gagatgaggt | 9780 |
| cagtgattaa atcacaacc catacttgc ataggaggag acacacaccc gtattctca | 9840 |
| ctggtagttc agttgagttt ctaatcttc gtgaccttgc tgctataatc agtaaagagt | 9900 |
| ctcaacatgt atattacctg acatggaaac tggttttgc gtattgtat gtcataagg | 9960 |
| ggaggtaat gacagagacc gctatgacta ttgatgttag gtatacagag cttctaggaa | 10020 |
| gagtcagata catgtggaaa ctgatagatg gtttcttcc tgcaactcggg aatccaactt | 10080 |
| atcaaattgt agccatgctg gagectctt cacttgcata cctgcagctg agggatataa | 10140 |
| cagttagaact cagagggtct ttcccttaacc actgctttac tgaaatacat gatgttctg | 10200 |
| acccaaaacgg gttttctgtat gaaggtactt atcatgagtt aattgaagct ctagattaca | 10260 |
| ttttcataac tgatgacata catctgacag gggagatccc ctcatcttc agaagttcg | 10320 |
| gccaccccgacttgaagca gtaacggctg ctgaaaaatgt taggaaatac atgaatcagc | 10380 |
| ctaaagtcat tgtgtatgag actctgtatgc aaggatcatgc catattttgtt ggaatcataa | 10440 |
| tcaacggctc tcgtgacagg cacggaggca gttggccacc gctgaccctc cccctgcata | 10500 |
| ctgcagacac aatccggaaat gctcaagctt caggtgaagg gttaacacat gagcagtg | 10560 |
| ttgataactg gagatctttt gctggagtga aatttggotc ctttatgcctt ctttagctgg | 10620 |
| atagtatgtc gacaatgtac ctaaaggaca aggacttgc tgctctccaa agggatggg | 10680 |
| attcagtttacccgaaagag ttccctgcgtt acgaccctcc caagggaaacc gggtcacgg | 10740 |

-continued

ggctttaga tgtttcctt aatgattcga gcttgaccc atatgatgtg ataatgtatg 10800
 ttgttaagtgg agcttacctc catgaccctg agttcaacct gtcttacagc ctgaaagaaa 10860
 aggagatcaa ggaaacaggt agacttttg ctaaaatgac ttacaaaatg agggcatgcc 10920
 aagtgattgc tgaaaatcta atctcaaacg ggattggcaa atattnaag gacaatggga 10980
 tggccaagga tgagcacgat ttgactaagg cactccacac tctagctgtc tcaggagtc 11040
 ccaaagatct caaaagaagt cacagggggg ggccagtctt aaaaacctac tccccgaagcc 11100
 cagtccacac aagtaccagg aacgtgagag cagcaaaagg gtttataggg ttccctcaag 11160
 taattcggca ggaccaagac actgtatc cggagaatat ggaagctac gagacagtca 11220
 gtgcatttat cacgactgat ctcaagaagt actgccttaa ttggagatata gacccatca 11280
 gettgtttgc acagaggcta aatgagattt acggattgc ctcatatccc cagtggctgc 11340
 ataagaggct tgagacctct gtcctgtatg taagtgaccc tcattgcccc cccgacctt 11400
 acgccccatat ccgttatata aaagtccccaa atgatcaaat cttcattaag taccctatgg 11460
 gaggtataga agggtattgt cagaagctgt ggaccatcag caccattccc tatctatacc 11520
 tggctgotta tgagagcgga gtaaggattt ctgcgttagt gcaagggac aatcagacca 11580
 tagccgtaac aaaaaggta cccagcacat ggccctacaa cottaagaaa cgggaagctg 11640
 ctagagtaac tagagattt tttgtatcc tttagcataag gctacatgat attggccatc 11700
 acctcaaggc aatgagaca attgtttcat cacattttt tgtctattca aaaggaatata 11760
 attatgatgg gctacttgtc tcccaatcac tcaagagcat cgcaagatgt gtattctgg 11820
 cagagactat agttgatgaa acaagggcag catgcgtaa tattgtacaa acaatggcta 11880
 aaagcatcga gagaggttat gaccgttacc ttgcataattc cctgaacgtc ctAAAAGTGA 11940
 tacagcaaatt tctgatctct cttggcttca caatcaattt aaccatgacc cgggatgtag 12000
 tcataccctt cctcacgaac aacgacctct taataaggat ggcactgtg cccgctct 12060
 ttggggggat gaattatctg aatatgacca ggctgtttgt cagaaacatc ggtgatcag 12120
 taacatcatc aattgtgtat ctcagagaa tgattctcg ctcactaatg cctgaagaga 12180
 ccctccatca agtaatgaca caacaacccg gggacttcc attcctagac tgggctagcg 12240
 acccttactc agcaaattt gtatgtgtcc agagcatcac tagactccctc aagaacataa 12300
 ctgcaagggtt tgcctgatc catagtccaa acccaatgtt aaaaggatata ttccatgatg 12360
 acagtaaaga agaggacgag ggactggcg cattcctcat ggacaggcat attatagtag 12420
 cttagggcagc tcatgaaatc ctggatcata gtgtcacagg ggcaagagag tctattgcag 12480
 gcatgctgga taccacaaaa ggcctgattc gagccagcat gaggaagggg gggtaacct 12540
 ctgcgtgtat aaccagattt tccaattatg actatgaaca attcagagca gggatgggtc 12600
 tattgacagg aagaaagaga aatgtcctca ttgacaaaga gtcatgttca gtgcagctgg 12660
 cgagagctct aagaagccat atgtgggcga ggctagctcg aggacggcctt atttacggcc 12720
 ttgaggtccc tcatgtacta gaatctatgc gaggccaccc tattcggcgt catgagacat 12780
 gtgtcatctg cgagatggta tcagtcaact acggatgtt tttgtcccc tcgggttgcc 12840
 aactggatga tattgacaaag gaaacatcat cttgagatgtt cccatataatt ggttctacca 12900
 ctgtatgagag aacagacatg aagcttgcct tcgtaaagac cccaaatgcg tccttgcgt 12960
 ctgctgttag aatagcaaca gtgtactcat gggcttacgg tcatgtatgtt agctcttgg 13020
 acgaaggctg gttgttggct aggcaaaaggg ccaatgtgag cctggaggag ctaagggtga 13080

-continued

| | | | | | | |
|-------------|-------------|--------------|-------------|--------------|------------|-------|
| tcaactccat | ctcaacttcg | actaatttag | cgcatacggtt | gagggatcgt | agcactcaag | 13140 |
| tgaaataatc | aggtaacatcc | cttgtccgag | tggcgaggta | taccacaatc | tccaacgaca | 13200 |
| atctctcatt | tgtcatatca | gataagaagg | ttgatactaa | ctttatatac | caacaaggaa | 13260 |
| tgcttctagg | gttgggtgtt | tttagaaacat | tgtttcgact | cgagaaaagat | accggatcat | 13320 |
| ctaacacggt | attacatctt | cacgtcgaaa | cgattgttg | cgtgatcccg | atgatagatc | 13380 |
| atcccaggat | acccagctcc | cgcaagctag | agctgagggc | agagctatgt | accaaccat | 13440 |
| tgatatatga | taatgcacct | ttaattgaca | gagatacaac | aaggctatac | acccagagcc | 13500 |
| ataggaggca | ccttggaa | tttggatcacat | ggtccacacc | ccaactatac | cacattttag | 13560 |
| ctaagtccac | agcactatct | atgattgacc | tggtaacaaa | atttggaaag | gaccatatac | 13620 |
| atgaaatttc | agctctcata | ggggatgacg | atataatag | tttcataact | gagttctgc | 13680 |
| tcatagagcc | aagattattc | actatctact | ttggccagtg | tgcggccatc | aattgggcat | 13740 |
| ttgatgtaca | ttatcataga | ccatcaggga | aatatcgat | gggtgagctg | ttgtcatcgt | 13800 |
| tcctttctag | aatgagcaaa | ggagtgttta | aggtgctgt | caatgtctta | agccacccaa | 13860 |
| agatctacaa | gaaattctgg | catttggta | ttatagagcc | tatccatgtt | ccttcacttg | 13920 |
| atgctcaaaa | cttgcacaca | actgtgtc | acatggttt | acatgttat | atgacctacc | 13980 |
| tcgacctgtt | gttgaatgaa | gagttttaga | agttcacatt | tctcttgcgt | gaaagcgacg | 14040 |
| aggatgttagt | acccggacaga | ttcgacaaca | tccaggcaaa | acacttatgt | gttctggcag | 14100 |
| atttgcactg | tcaaccaggg | acctgcccac | caattcgagg | tctaagacgc | gtagagaaat | 14160 |
| gtgcagttct | aaccgaccat | atcaaggcag | aggcttaggt | atctccagca | ggatcttcgt | 14220 |
| ggaacataaaa | tccaattatt | gtagaccatt | actcatgctc | tctgacttat | ctccggcgag | 14280 |
| gatcgatcaa | acagataaga | tttagagttt | atccaggatt | cattttcgac | gccctcgctg | 14340 |
| aggttaatgt | cagtcagcc | aagatcgcc | gcaacaacat | ctcaaataatg | agcatcaagg | 14400 |
| cttcagacc | cccacacgt | gatgtgtca | aattgctcaa | agatatacAAC | acaagcaac | 14460 |
| acaatcttcc | catttcaggg | ggcaatctcg | ccaattatga | aatccatgtt | ttccgcagaa | 14520 |
| tcgggttgaa | ctcatctgct | tgctacaaag | ctgttgagat | atcaacat | attaggagat | 14580 |
| gccttgagcc | aggggaggac | ggcttgcgt | tgggtgagg | atcgggttcc | atgttgatca | 14640 |
| cttataagga | gatactaaaa | ctaaacaatgt | gcttctataa | tagtgggg | tccgcatt | 14700 |
| ctagatctgg | tcaaaggaa | ttagcaccct | atccctccg | agttggccct | gtcgaacaca | 14760 |
| gaatgggagt | aggtatattt | gtcaaagtgc | tcttaacgg | gaggccccaa | gtcacgtgg | 14820 |
| taggcagtgt | agattgttc | aatttcatag | ttagtaat | ccctacctt | agtgtgggt | 14880 |
| ttatccattc | agatatacag | accttgcata | acaaagatac | tatagagaag | ctagagaaat | 14940 |
| tggcagccat | tttatcgat | gctctgc | tggcaaaat | aggatcaata | ctggtgatta | 15000 |
| agcttatgcc | tttcagccgg | gattttgttc | agggatttt | aagttatgt | gggtccatt | 15060 |
| atagagaagt | gaaccttgt | taccctagat | acagcaactt | catactact | gaatcttatt | 15120 |
| tggttatgac | agatctcaag | gctaaacggc | taatgaatcc | tggaaaaggatt | aagcagcaga | 15180 |
| taattgaatc | atctgtgagg | acttcacctg | gacttatagg | tcacatccta | tccattaac | 15240 |
| aactaagctg | catacaagca | attgtggag | acgcagttag | tagaggtgt | atcaatccta | 15300 |
| ctctgaaaaa | acttacacct | atagagcagg | tgctgatcaa | ttgcgggttg | gcaattaac | 15360 |
| gacctaagct | gtgcaaagaa | ttgatccacc | atgatgtgc | ctcaggccaa | gatggatgc | 15420 |
| ttaattctat | actcatcctc | tacaggag | tggcaagatt | caaagacaac | caaagaagtc | 15480 |

-continued

```
aacaaggat gttccacgct taccggat tggtaagtag caggcaacga gaacttatat 15540
ctaggatcac ccgcaaattt tgggggcaca ttcttcgttta ctccgggaac agaaaagtga 15600
taaataagtt tatccagaat ctcaagtccg gctatctgat actagactta caccagaata 15660
tcttcgttaa gaatctatcc aagtcagaga aacagattat tatgacgggg ggtttgaac 15720
gtgagtgggt ttttaaggta acagtcaagg agaccaaaga atggtataag ttatcgat 15780
acagtgcctt gattaaggac taagcggcc 15809
```

```
<210> SEQ ID NO 10
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E
      genome
```

<400> SEQUENCE: 10

gcggccgcac caaac 15

```
<210> SEQ ID NO 11
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E
      genome
```

<400> SEQUENCE: 11

cctgaccgcg gatgc 15

```
<210> SEQ ID NO 12
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E
      genome
```

<400> SEQUENCE: 12

acctcgcatc cgccgg 15

```
<210> SEQ ID NO 13
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E
      genome
```

<400> SEQUENCE: 13

cctccagagt aatcgattaa gg 22

```
<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E
      genome
```

<400> SEQUENCE: 14

aatcgattac tctggaggag cag 23

-continued

<210> SEQ ID NO 15
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 15

cttgcaccct aagtttaat taactac

27

<210> SEQ ID NO 16
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 16

gaacaatac ggttagttaat taaaac

26

<210> SEQ ID NO 17
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 17

tgagggactc gagcatactc

20

<210> SEQ ID NO 18
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 18

ataagatagt agccatcctg gagtat

26

<210> SEQ ID NO 19
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 19

gtagggccat gtgctggg

18

<210> SEQ ID NO 20
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 20

catagccgta acaaaaaggg tac

23

<210> SEQ ID NO 21

-continued

<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 21

gagccatcaag tgaaggacca tg

22

<210> SEQ ID NO 22
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 22

gcattgtggt attatacgac ctatc

25

<210> SEQ ID NO 23
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer for synthesis of cDNA of entire MV-E genome

<400> SEQUENCE: 23

cggtttaaac cagacaaagc tg

22

<210> SEQ ID NO 24
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Forward (N) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 24

gctagcatgg ccacactttt aagg

24

<210> SEQ ID NO 25
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse (N) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 25

gcggccgcct agtctagaag att

23

<210> SEQ ID NO 26
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Forward (P) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 26

gctagcatgg cagaagagca gg

22

<210> SEQ ID NO 27
<211> LENGTH: 24

-continued

<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse (P) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 27

gccccgcct acttcattat tata

24

<210> SEQ ID NO 28
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Forward (L) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 28

gttagcatgg actcgctatc tgtcaac

27

<210> SEQ ID NO 29
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse (L) random hexamer for RNA preparation from purified MV-E virus

<400> SEQUENCE: 29

gccccgcct agtccttaat cag

23

<210> SEQ ID NO 30
<211> LENGTH: 130
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: pentameric IRES elememt flanked by a site for EagI at 5' and first 10 nucleotides of L protein at 3'

<400> SEQUENCE: 30

ggccgttctg acatccggcg ggttctgac atccggcggtt ttctgacat ccggcggtt 60

tctgacatcc ggccggtttc tgacatccgg cgggtgactc acaacggatc caacagacat 120

atggactcgc 130

<210> SEQ ID NO 31
<211> LENGTH: 91
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: porcine teschovirus 2A peptide flanked on the 5'end by the codons immediately before stop codon of MV N protein and on the 3'end by the first few codons of MV P protein

<400> SEQUENCE: 31

atttctaga cggctccgga gccacgaact tctctctgtt aaagcaagca ggagacgtgg 60

aagaaaaacc cggtccatg gcagaagagc a 91

67

The invention claimed is:

1. A two plasmid system for producing recombinant proteins using RNA dependent RNA Polymerase of non-segmented negative strand RNA viruses comprising:

- a. one cloning plasmid that expresses the recombinant proteins, comprising a manipulatable replicon having at least two multiple cloning sites for inserting genes encoding the recombinant proteins, and
- b. one helper plasmid that expresses N, P, and L proteins, comprising N, P, and L genes expressing the N, P, and L proteins, respectively;

without the help of a helper vaccinia virus or exogenous RNA polymerase, wherein the manipulatable replicon sequence is as set forth in SEQ ID NO: 1.

2. The two plasmid system as claimed in claim **1**, wherein the helper plasmid is selected from the group consisting of SEQ ID NO: 8, and SEQ ID NO: 9.

3. The two plasmid system as claimed in claim **1**, wherein the cloning plasmid is selected from the group consisting of SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, and SEQ ID NO: 7.

4. A two plasmid system for producing recombinant proteins using RNA dependent RNA Polymerase of non-segmented negative strand RNA viruses comprising:

- a. one cloning plasmid that expresses the recombinant proteins, comprising a manipulatable replicon having at least two multiple cloning sites for inserting genes encoding the recombinant proteins, and
- b. one helper plasmid that expresses N, P, and L proteins, comprising N, P, and L genes expressing the N, P, and L proteins, respectively;

68

without the help of a helper vaccinia virus or exogenous RNA polymerase, wherein the manipulatable replicon is flanked by ribozymes, and said manipulatable replicon sequence is as set forth in SEQ ID NO: 2.

5. The two plasmid system as claimed in claim **4**, wherein the helper plasmid is selected from the group consisting of SEQ ID NO: 8, and SEQ ID NO: 9.

6. The two plasmid system as claimed in claim **4**, wherein the cloning plasmid is selected from the group consisting of SEQ ID NO: 4, and SEQ ID NO: 7.

7. A two plasmid system for producing recombinant proteins using RNA dependent RNA Polymerase of non-segmented negative strand RNA viruses comprising:

- a. one cloning plasmid that expresses the recombinant proteins, comprising a manipulatable replicon having at least two multiple cloning sites for inserting genes encoding the recombinant proteins, and
- b. one helper plasmid that expresses N, P, and L proteins, comprising N, P, and L genes expressing the N, P, and L proteins, respectively;

without the help of a helper vaccinia virus or exogenous RNA polymerase, wherein the manipulatable replicon is flanked by RNA polymerase I promoter and RNA polymerase I terminator, and said manipulatable replicon sequence is as set forth in SEQ ID NO: 3.

8. The two plasmid system as claimed in claim **7**, wherein the helper plasmid is selected from the group consisting of SEQ ID NO: 8, and SEQ ID NO: 9.

9. The two plasmid system as claimed in claim **7**, wherein the cloning plasmid is selected from the group consisting of SEQ ID NO: 5, and SEQ ID NO: 6.

* * * * *